Structural Studies on the eIF4A-eIF4G Interaction in Translation Initiation

Dissertation Advisor: Professor Gerhard Wagner Katherine Ann Edmonds

Structural studies on the eIF4A-eIF4G interaction in translation initiation

Abstract

Protein synthesis is an important cellular process, and the RNA helicase eIF4A plays a vital role in unwinding messenger RNA and scanning during translation initiation. eIF4A has little activity in isolation, but is modulated by other initiation factors such as eIF4G and eIF4H. In this thesis, we explore how these proteins come together to form a functional unwinding complex. We begin with the NMR solution structure of a single domain from this complex, eIF4G HEAT2. We then map interactions involving HEAT2 and its binding partners, as well as those involving the N-terminal domain of eIF4A. We use this information first to construct a structure of the two-domain complex of HEAT2 and eIF4A-NTD, and expand this work toward the structure of the 70kDa, three-domain complex of HEAT2 with full-length eIF4A. Finally, we incorporate eIF4H and another domain of eIF4G to model the entire functional complex, and explore how interactions between domains rearrange upon binding, hydrolysis, and release of ATP. These results give us a better understanding of how eIF4G modulates eIF4A helicase activity. Moreover, the domain organization of the complex allows us to construct a more compelling model to explain how eIF4A facilitates preinitiation complex scanning along a messenger RNA.

Contents

Author List

This thesis is the work of Katherine Edmonds. The following authors also contributed to work presented in this thesis:

Much of the work presented in Chapters 4 and 5 was performed in collaboration with, and under the supervision of, Assen Marintchev. He recorded the spectra shown in the right panel of Figure 4.1. He also recorded earlier versions of the spectra shown in the left panel of the same figure. Assen performed the fluorescence anisotropy assay shown in Figure 4.14. Results from his work are shown in Figures 5.6, 5.7, 5.8, 5.9, 5.10, and 5.13.

Chikako Suzuki recorded the spectra that were used to make the chart in the bottom panel of Figure 4.13.

The FRET experiments described in Section 5.2 were part of a collaboration with Thomas-Otavio Peulen, Markus Richert, and Claus Seidel.

Listing of figures

Acknowledgments

I WOULD LIKE TO EXPRESS MY GRATITUDE to Gerhard, not just for how he has mentored me, but also for establishing such a rich environment for learning NMR and protein production. So many members of the Wagner lab have influenced my growth as a scientist in so many ways, it would be impossible to name them all.

First, I am thankful to Mallika, for stepping in and emphasizing the importance of the "backup project", and in doing so, for steering me toward projects that I would find more stimulating. Monika and Chikako took me under their wings and taught me everything I needed as a foundation for challenging protein purification work. Assen played an important role organizing, directing, and mentoring the translation initiation subgroup. Thanks to Alex for pushing me in NMR and especially for providing an audience for so many of the scripts that I wrote to facilitate NMR data analysis.

I could not have done this work without all the help I have gotten from Maura and from Michele in navigating the system.

Finally, I appreciate the support I received from my family throughout this process, and especially from Sam, who had to put up with me on a daily basis.

1

Introduction

1.1 Translation and protein synthesis

Translation of messenger RNA (mRNA) into proteins is one of the most important cellular processes. Protein synthesis begins with the assembly of the small ribosomal subunit with several initiation factors and the initiator transfer RNA into a 43S preinitiation complex. The next step, in which the preinitiation complex binds to the mRNA and locates the start codon, differs considerably between prokaryotes and eukaryotes. Start codon recognition then leads to GTP hydrolysis and joining of the large ribosomal subunit. Assembly of the

amino acids according to the mRNA template can finally commence after another round of GTP hydrolysis. The translation process then goes through the elongation phase, and ultimately terminates when a stop codon is reached, and ribosomal subunits as well as initiation, elongation, and termination factors are recycled for further rounds of protein synthesis.

1.2 Translational control and the role of initiation factors

Translational control is important for regulating cellular protein levels. Controlling expression rates at the point of translation rather than transcription provides the advantage of faster response to environmental changes. Regulation occurs at many steps of the translation process, but initiation is typically the rate-limiting step of protein synthesis [47], making regulation at this phase particularly effective. It is a complex, multi-step process to bring together the small and large ribosomal subunits, along with many initiation factors and the initiator transfer RNA at the start codon of an mRNA. In this thesis, we focus on the middle phase of initiation, which includes the uniquely eukaryotic processes of mRNA activation and scanning through the 5'-untranslated region (5'-UTR) to find the start codon.

Figure 1.1 outlines the process of recruiting the small ribosomal subunit to the messenger RNA and scanning to the AUG start codon in eukaryotic cap-dependent translation initiation. mRNAs in eukaryotes typically have a 7-methyl guanosine triphosphate (m^7GTP) cap at the 5' end, which significantly enhances translation efficiency, compared to initiation on uncapped mRNA. In the first step, the eukaryotic Initiation Factor 4F (eIF4F) complex binds to the 5' end of the mRNA, through one of its component proteins, eIF4E, which specifically recognizes the m^7GTP cap. If eIF4E was not already in an eIF4F complex, the cap binding event significantly increases eIF4E affinity for eIF4G [23], bringing this factor

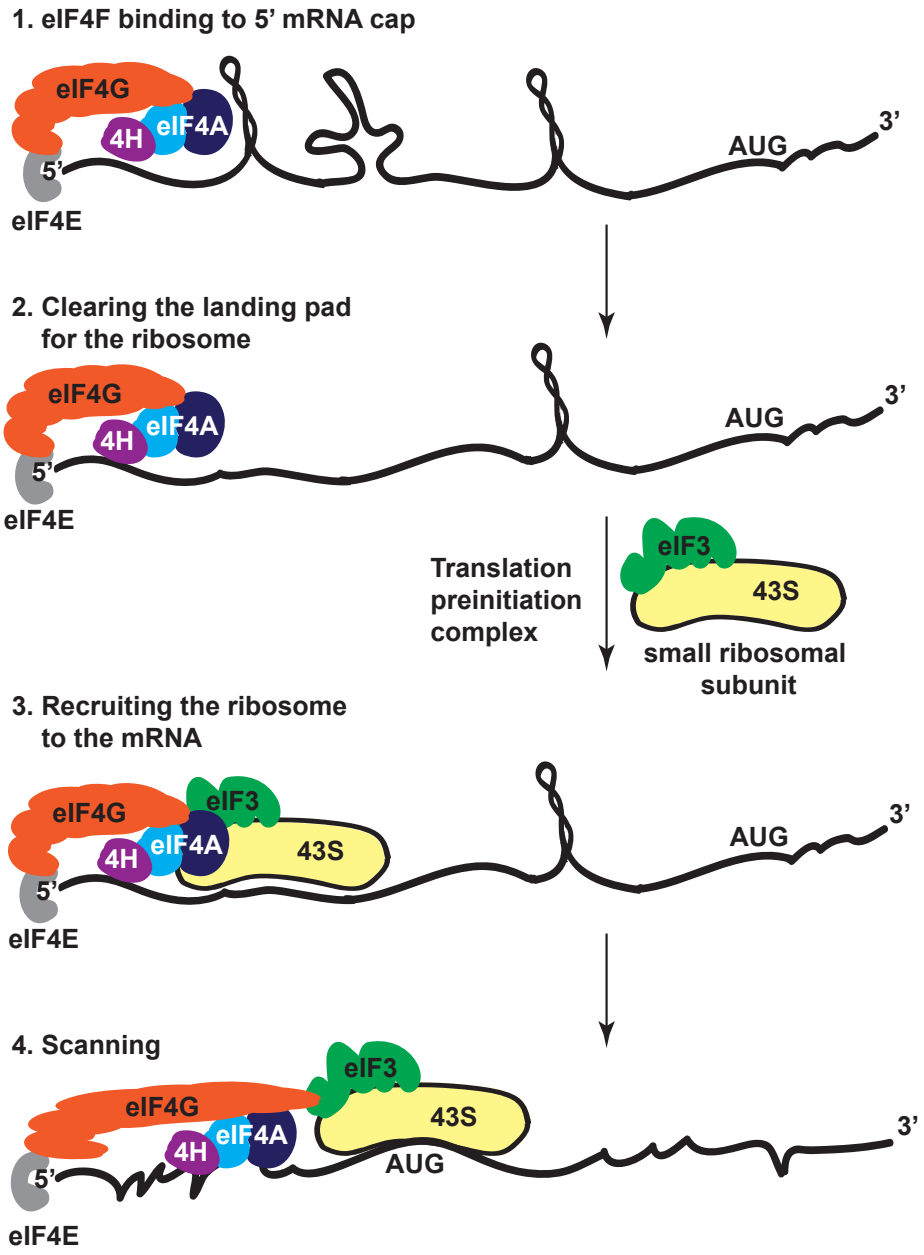

Figure 1.1: Overview of the role of proteins in the eIF4 family of initiation factors in cap-dependent translation. These proteins are responsible for recruiting the preinitiation complex including the small ribosomal subunit to the 5' end of the mRNA for scanning to the start codon. eIF4E recognizes the 5' cap of the mRNA, eIF4A is responsible for unwinding of mRNA secondary structure to "clear a landing pad" for the small ribosomal subunit, eIF4H enhances eIF4A activity, and eIF4G recruits the 43S preinitiation complex through its interaction with eIF3, while also serving as the scaffolding protein that brings eIF4E and eIF4A together.

and eIF4A, which are also components of eIF4F, to the 5' end of the mRNA. eIF4G is alarge scaffolding protein which serves as the central component of eIF4F, and which harbors additional binding sites for other initiation factors, including the RNA helicase eIF4A, which completes the eIF4F complex. eIF4A also binds to the RNA binding protein eIF4H, which stimulates its helicase activity.

Once these factors are assembled on the mRNA, they are responsible for the second step shown in Figure 1.1, unwinding secondary structure in the 5'UTR, to clear a "landing pad" for the 43S preinitiation complex. In the third step shown in Figure 1.1, the 43S preinitiation complex is recruited to the mRNA. The preinitiation complex contains the small ribosomal subunit and many other translation initiation factors, including the multisubunit initiation factor eIF3. The recruitment process is believed to be mediated by an interaction between eIF4G and eIF3. Finally, the preinitiation complex scans along the mRNA to the AUG start codon. Here, the initiation factors are released, and the 60S large ribosomal subunit binds to form the 80S ribosome that proceeds with protein synthesis.

Changes to the translation initiation process favor some messenger RNAs over others. Strong mRNAs have short 5' UTRs with little secondary structure, and are easily translated even under suboptimal conditions. Weak mRNAs, on the other hand, have long, complex leader sequences with structure that impedes the progress of scanning ribosomes. eIF4A and eIF4F are crucial for unwinding the long, highly structured 5'UTRs commonly found in weak mRNAs such as those commonly found encoding growth factors and oncogenes [61].

Figure 1.2 outlines some of the initiation factors mentioned in Figure 1.1 in more detail, showing key domains and known structures that are important for this work. This figure also shows a color code that is used as much as possible throughout the figures in this thesis. For example, the N-terminal domain of eIF4A is consistently depicted in dark blue, and the C-terminal domain in light blue, while the HEAT2 domain of eIF4G is mostly shown

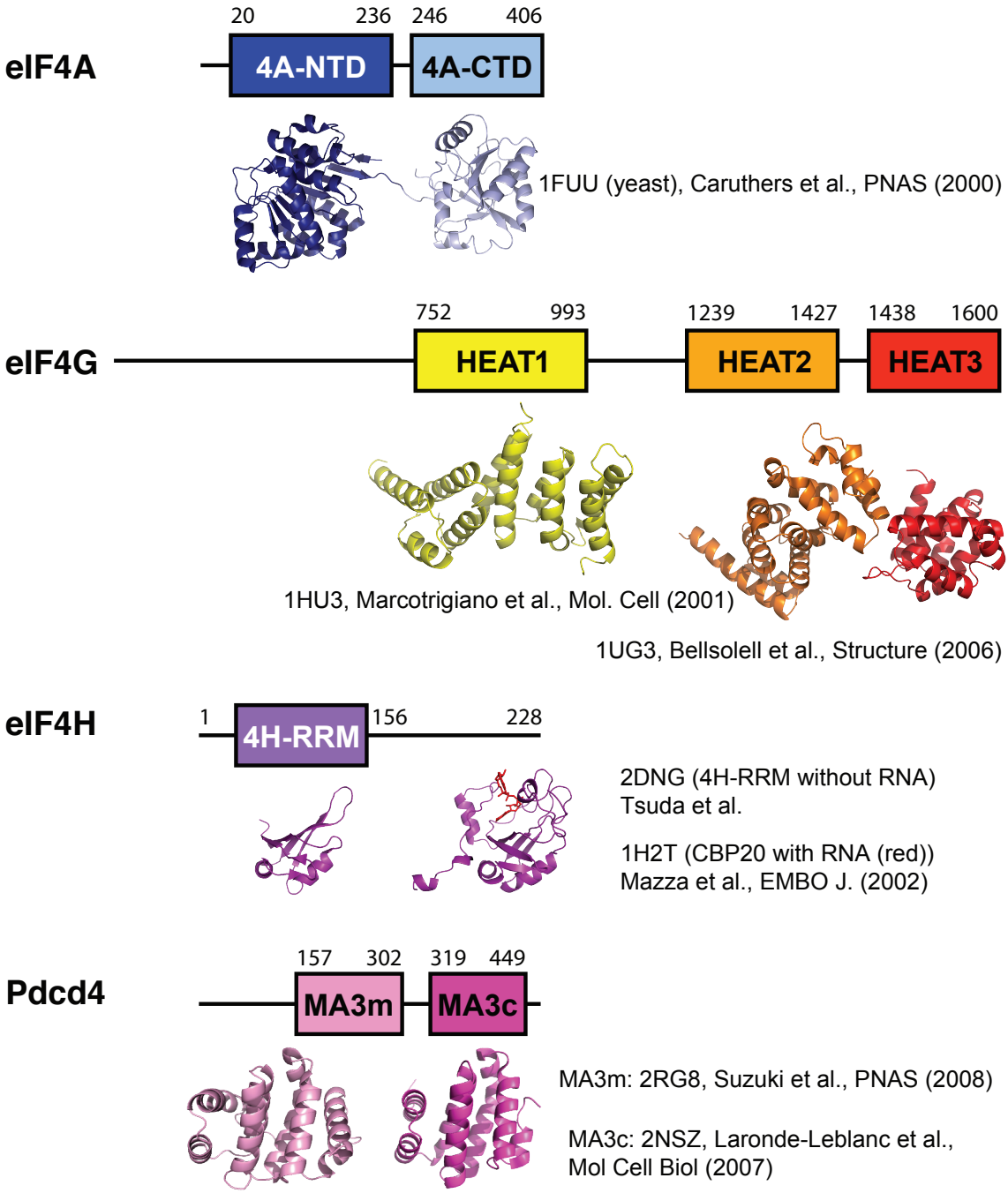

Figure 1.2: Overview of the translation initiation factors and inhibitors used in this work, including structures of domains or homologous domains [5, 8, 35, 40, 43, 60].

Proteins or initiation factors	organism	pdb code	reference
eIF4A	yeast	1FUU	[8]
eIF4A NTD	human	2G9N	[53]
VASA (DEAD-box RNA helicase)	drosophila	2DB3	[56]
eIF4AIII in exon junction core complex	human	2HYI	[2]
eIF4A + eIF4G HEAT1	yeast	2VSO	[52]
eIF4G HEAT2-HEAT3	human	1UG3	[5]
eIF4A + Pdcd4	mouse / human	3EIQ	[37]
eIF4A + Pdcd4	human	2ZU6	[9]
MjDEAD	M. jannaschii	1HV8	[59]
CBP80 + CBP20	human	1H2T	[43]

Table 1.1: Domains and structures discussed in this work.

in orange. These proteins are described in more detail in the following sections, and the structures shown are listed in Table 1.1 along with other structures mentioned in this text.

1.3 Eukaryotic Initiation Factor 4A (eIF4A)

The eukaryotic initiation factor 4A (eIF4A) is the most abundant eukaryotic translation initiation factor [12], and is the prototypical member of the DEAD-box family of ATP-dependent RNA helicases. This family obtains its name from the conserved sequence motif Asp-Glu-Ala-Asp which is believed to be responsible for coordinating a magnesium ion which in turn activates a water molecule to hydrolyze the terminal phosphate from ATP. In addition to having ATP-dependent unwinding activity, they also exhibit RNA-dependent ATPase activity. DEAD-box proteins typically consist of a core of two recA-like domains which contain a set of conserved motifs responsible for binding ATP and RNA, along with flanking domains with auxiliary functions that may serve to enhance substrate binding or have other roles.

eIF4A consists only of the two core recA-like domains, lacking the auxiliary domains that many other DEAD-box helicases have, although other initiation factors that interact with

eIF4A may play some of the same roles. Figure 1.3 outlines the motifs commonly found in DEAD-box proteins, and highlights their positions on the structure of the N-terminal domain (eIF4A-NTD, PDBID 2G9N, top left panel). The top right panel of Figure 1.3 shows these motifs on a model of full-length human eIF4A that is based on the crystal structure of the related protein eIF4AIII, which is involved in RNA processing, but not in translation initiation [2]. The RNA and ATP analog AMPPNP shown as sticks in these models are taken directly from the eIF4AIII crystal structure.

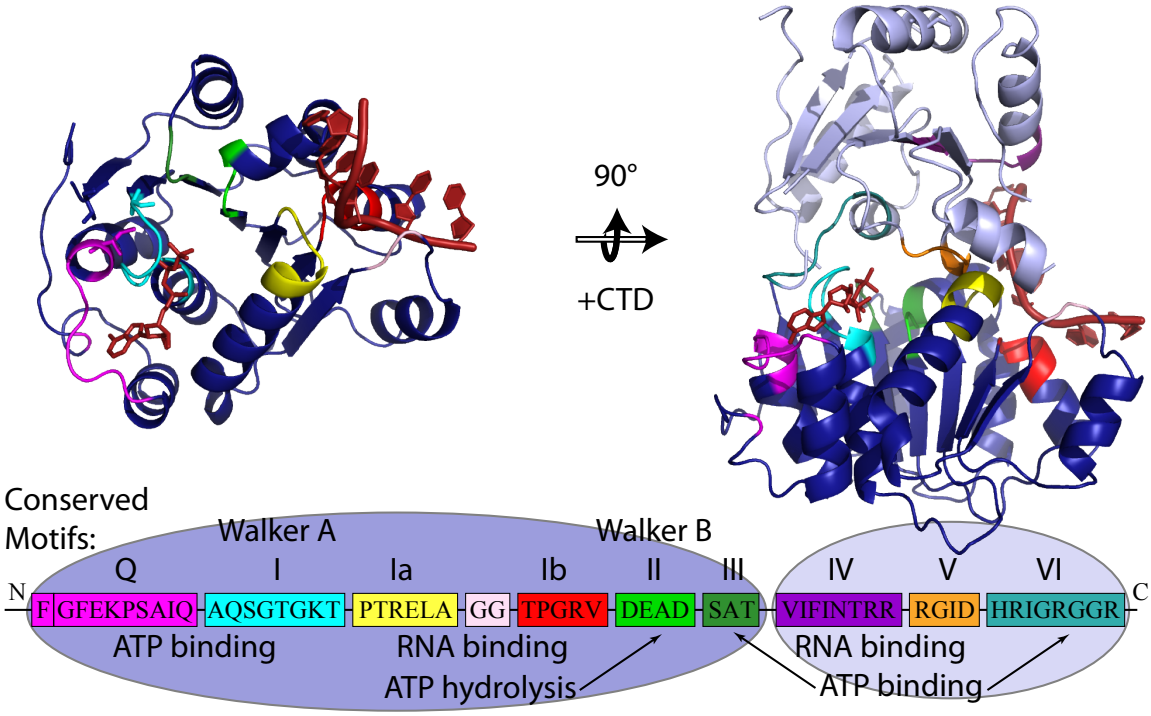

Figure 1.3: Conserved DEAD-box helicase motifs in eIF4A. The crystal structure of human eIF4A-NTD (2G9N) is shown on the upper left, colored according to conserved regions involved in ATPase or helicase activity. This structure is accompanied by a model of human eIF4A-CTD, also with color coded motifs matching the scheme shown below. The domains are positioned with RNA and ATP modeled as they appear in the crystal structure of eIF4AIII (2HYI) [2].

The conserved DEAD-box motifs are listed in the bottom panel of Figure 1.3, in the order that they appear in the eIF4A sequence. The N-terminal domain, shown here and in

much of the thesis in dark blue, contains most of the motifs responsible for ATP binding, including the well-known motifs I and II, also known as Walker A and Walker B which are responsible for coordination and hydrolysis of the phosphates in ATP. eIF4A-NTD also contains the Q motif [62], which binds the adenine of ATP. The remaining motifs on the NTD, including Ia, Ib, GG, and SAT, have been implicated in RNA binding or in helicase activity.

The smaller C-terminal domain contains motifs IV and V, which are implicated in RNA binding, and motif VI, which is involved in binding ATP. Although these conserved motifs are distributed throughout the full sequence of eIF4A, they are all localized to the cleft between the two domains in the modeled structure. The motifs labeled as responsible for RNA binding can all be seen to cluster on the right hand side of this cleft in the Figure: motif IV, colored purple, is directly above the modeled RNA, while the orange motif V and yellow Ia are immediately to the left, Ib is below, in red, and the pink GG motif is directly in front of the RNA from this perspective. Meanwhile, the Q motif and motifs I, II, and VI surround the AMPPNP on the left side of the cleft.

There are no crystal structures of full-length human eIF4A in isolation, but there is a crystal structure of the N-terminal domain (2G9N), and of full-length eIF4A from yeast (1FUU) [8]. The yeast structure shows eIF4A in an extended conformation (shown in the left panel of Figure 1.4), which is not likely to be an active form, since the motifs responsible for ATP binding, RNA binding, and helicase activity are so far apart. These domains must surely come together to form continuous binding surfaces for ATP and RNA in order to function as an active helicase, and ATP and RNA may even serve to bring the domains together into an active form. Therefore, we use other structures of DEAD-box helicases, which have been crystallized in closed forms in the presence of RNA and nucleotide, to model what we believe to be the active form of eIF4A. We have already mentioned the structure

of human eIF4AIII from the exon junction core complex (PDB 2HYI) [2], which is the primary source of the closed form models used in this thesis, shown in the right panel of Figure 1.4. This approach is validated by the existence of several other examples of DEAD-box helicases crystallized in closed conformations with RNA and nucleotide, including the drosophila helicase VASA [56], or MjDEAD from Methanococcus jannaschii [59]. More recently, full-length human eIF4A has been crystallized in the presence of the translation repressor protein Pdcd4 [9, 37], and we discuss the use of these structures for modeling active conformations of eIF4A later in this thesis (Section 5.3.5).

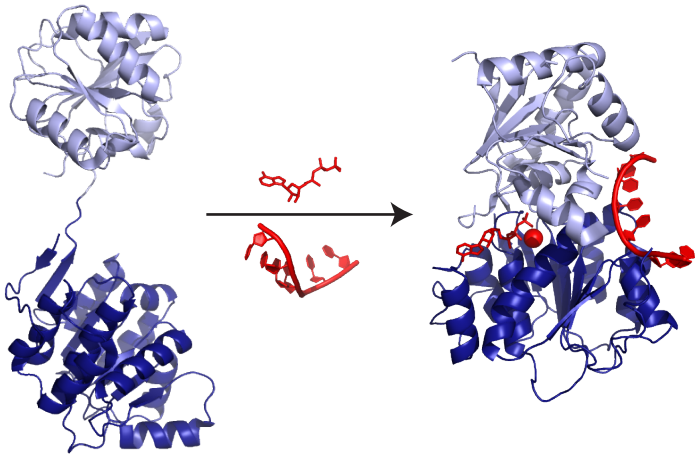

Figure 1.4: Crystal structures of yeast eIF4A (left, 1FUU) and human eIF4AIII in a larger complex with RNA and an ATP analog inform a basic model of eIF4A activity. Here, the human eIF4AI domains (from 2ZU6 [9] and 2G9N) are modeled in a closed and presumably active conformation capable of binding ATP and RNA (right), as they appear in the structure of eIF4AIII (2HYI) [2]. The open yeast structure represents the inactive conformation of eIF4A, in the absence of ATP and RNA.

eIF4A exhibits very low helicase activity on its own, but its unwinding efficiency is enhanced by interactions with other initiation factors such as eIF4G, eIF4H, and eIF4B [49]. One mechanism which has been proposed for enhancing eIF4A activity is to promote the closed, presumably more active form of the helicase, as shown in the right panel of Figure 1.4, at the expense of more open states such as the one shown in the left panel. The HEAT1

domain of eIF4G has been proposed to enhance eIF4A activity by promoting a relatively closed form of the helicase, through simultaneous interactions with both eIF4A-NTD and eIF4A-CTD [45].

1.4 EUKARYOTIC INITIATION FACTOR 4G (EIF4G)

The translation initiation factor eIF4G is a large scaffolding protein at the center of the eIF4F complex, bridging between the 5'-cap binding protein eIF4E and the RNA helicase eIF4A, as well as harboring binding sites for a variety of other proteins involved in translation initiation. eIF4G has historically been conceptually divided into thirds, according to picornaviral protease cleavage products [34]: the N-terminal third contains the eIF4E binding site, as well as a binding site for the Poly-A binding protein that interacts with the 3' end of an mRNA. The middle third of eIF4G interacts with the multisubunit initiation factor eIF3, which binds directly to the ribosome, and the C-terminal third contains a binding site for the kinase MNK1. Mutational and deletional analyses have shown that two distinct eIF4A binding regions are present in the C-terminal two thirds of mammalian eIF4G, one in the middle segment, and one in the C-terminal segment [29], although yeast lack the latter of these two sites.

These two eIF4A binding regions fall mostly within the domains HEAT1 and HEAT2, which are highlighted in yellow and orange, respectively, in Figure 1.2. These domains are HEAT-repeat domains, composed of several tandem HEAT repeats, each of which consists of a pair of α-helices connected by a short linker. HEAT-repeat domains, also called HEAT domains, are a common motif for protein-protein interactions. The first HEAT domain, HEAT1, stimulates the helicase activity of eIF4A. The role of HEAT2 is less clear. It is not required for translation initiation; it is not even present in yeast eIF4G. Rather, it plays some modulatory role and appears to be somewhat inhibitory (Korneeva et al., 2005).

A core region of eIF4G containing the eIF4E and eIF3 binding sites, as well as the first of the two eIF4A binding regions has been identified as crucial for supporting cap-dependent mRNA translation. The C-terminal eIF4A binding region is not required for translation, but it has been shown to improve translational efficiency in rabbit reticulocyte lysate, since a truncation mutant lacking HEAT2 and a HEAT2 point mutant deficient in eIF4A binding both fail to stimulate translation to the same level as wildtype eIF4G [44].

1.5 eIF4H and eIF4B enhance eIF4A helicase activity

eIF4H is a small RNA-binding protein with considerable homology to the much larger protein eIF4B. eIF4H contains an RNA recognition motif (RRM), along with flexible N- and C-terminal regions. eIF4H is thought to enhance eIF4A helicase activity by acting as a tether to the RNA substrate [49]. Although eIF4H does not appear to enhance eIF4A helicase activity as strongly as eIF4B [46], we have chosen to focus on eIF4H in our studies because its small size makes it more suitable for NMR spectroscopy.

1.6 Pdcd4 contains two MA3 domains that resemble HEAT2

Pdcd4 is a tumor suppressor with two MA3 domains, each of which is homologous and structurally similar to the HEAT2 domain of eIF4G [35, 60]. Figure 1.5 shows the alignment between eIF4G HEAT2 and the middle and C-terminal domains of Pdcd4, called MA3m and MA3c, respectively. Pdcd4 has been shown to inhibit cap-dependent translation initiation through its interaction with eIF4A [69]. Several mutations in Pdcd4 have been shown to interfere with both its roles of translational inhibition, and eIF4A binding. Analogous mutations in the HEAT2 domain of eIF4G also interfere with eIF4A binding to eIF4G [70], suggesting that the similarities between HEAT2 and Pdcd4 may extend beyond the structure, to their functional roles in interacting with eIF4A.

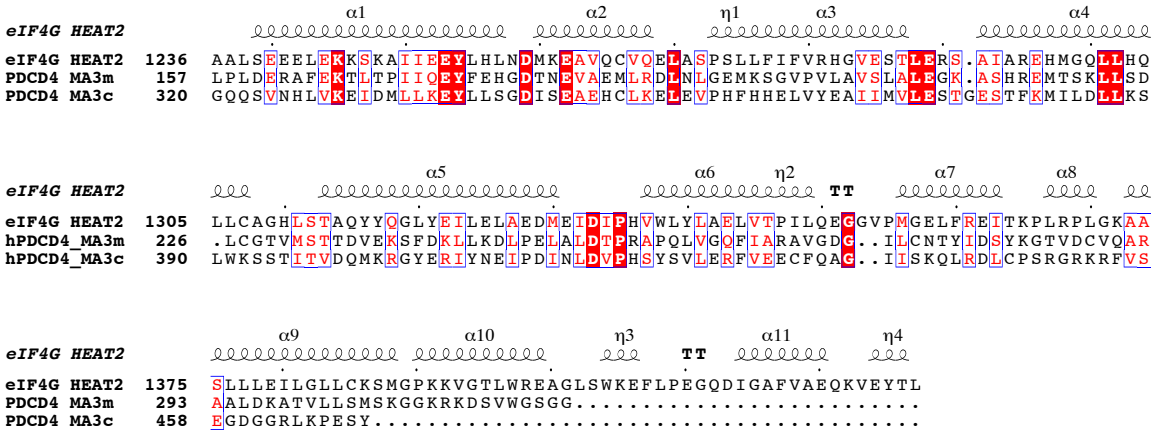

Figure 1.5: Sequence alignment of human eIF4G HEAT2 with both MA3 domains of human Pdcd4. The sequence and secondary structure alignment was prepared using UniProt [10] and ESPript [22]

1.7 Starting model for the assembly of eIF4 proteins

It is well-known that eIF4A functions as a helicase in translation initiation primarily in the context of larger protein complexes. In particular, eIF4A can be purified as a component of eIF4F, which is held together by the strong interactions eIF4G makes with the cap-binding protein eIF4E and with eIF4A. Additional factors such as eIF4B and eIF4H considerably improve the unwinding efficiency and processivity of eIF4A [49], and have been shown to bind directly to eIF4A [15]. Our goal is to gain a better understanding of how these factors fit together to form a functional complex, building up from pairwise interactions to larger subcomplexes. Many of the interactions are already familiar from mutagenesis studies, combined with pull-down experiments or functional assays.

Figure 1.6 illustrates our initial hypothesis for the configuration of the active complex in which eIF4A is able to unwind the secondary structure in the 5' UTR of mRNA. The left panel shows how we model the active conformation of full-length eIF4A, with the two domains coming together to form continuous binding surfaces for RNA and ATP, as well as any other factors that interact with both domains. This modeled structure is based on the

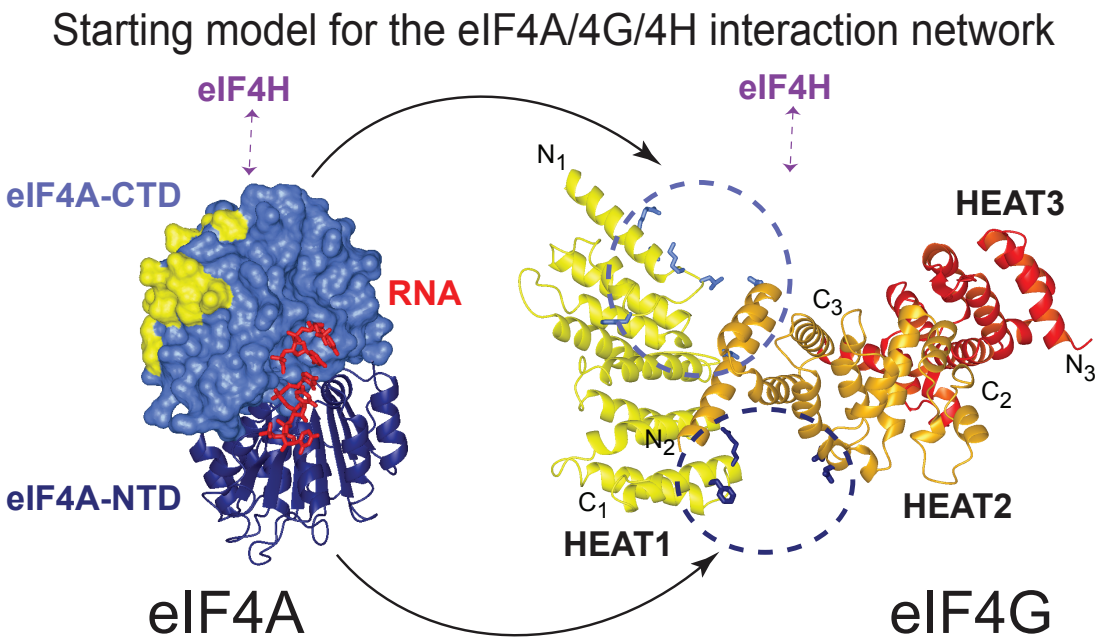

Figure 1.6: Starting model for the assembly of eIF4A, eIF4G, and eIF4H proteins. The left panel shows eIF4AI domains modeled in a closed conformation based on the structure of eIF4AIII with RNA and an ATP analog [2], with the previously determined HEAT1 binding surface painted yellow [45]. The right panel shows eIF4G domains oriented as in the structure of the homologous CBP80 [41, 43]. The expected position of eIF4H is marked in both panels. Sites of mutations in eIF4G reported to affect eIF4A binding are shown as sticks, colored light blue in the top cluster (circled), and dark blue in the bottom cluster (circled). Arrows point from each domain of eIF4A to its expected location in the larger complex. Figure adapted from Marintchev et al. 2009 [42].

crystal structure of eIF4AIII in an exon junction core complex, containing RNA and a non-hydrolyzable ATP analog [2]. The previously mapped binding site for the HEAT1 domain on the C-terminal domain of eIF4A is colored yellow [45]. The right panel of Figure 1.6 shows the starting model of how the eIF4G domains are oriented with respect to each other, based on the crystal structure of the homologous CBP80, which also contains three HEAT domains [41, 43]. The same structure also includes CBP20, which is homologous to eIF4H, and suggests a location for eIF4H above the HEAT domains in this picture.

Many mutations in eIF4G have been described which impact binding to eIF4A [70], some of which are shown as light blue and dark blue sticks in Figure 1.6. Although these mutations

appear to be dispersed among a wide range of positions in the eIF4G sequence, they appear to form two distinct clusters in space when HEAT1 and HEAT2 are arranged in this way. In particular, this structural arrangement places residues in the first interhelical linker of HEAT2 near residues in the first interhelical linker of HEAT1, raising the possibility that mutations to these residues might affect binding to the same domain of eIF4A. Likewise, this arrangement of domains places known mutants in the third HEAT repeat of HEAT2 close to residues near the C-terminus of the HEAT1 domain which are known to affect eIF4A binding. We hypothesize that this latter group of residues is important for binding to eIF4A-NTD, while the former group is responsible for binding to eIF4A-CTD.

1.8 Structure of this thesis

This thesis presents an investigation of the supramolecular assembly which surrounds eIF4A and regulates its helicase activity. We begin with Chapter 2, which describes the materials and methods used throughout this thesis, followed by three chapters exploring progressively larger components of the eIF4A/eIF4G/eIF4H assembly.

Chapter 3 is devoted to determining the solution structure of the HEAT2 domain of eIF4G by nuclear magnetic resonance spectroscopy (NMR). A crystal structure of HEAT2 has previously been solved [5], and we find that the solution structure is very similar to the crystal structure, with the exception of an unusually long N-terminal helix, which appears to be a packing artifact in the crystal structure. Some of the work presented in this chapter has previously been published in Gal et al. 2011 [18].

In Chapter 4, we explore and map some of the pairwise interactions between domains in our model of the active unwinding complex (Figure 1.6). Wherever possible, we use chemical shift perturbation assays to map the binding interfaces of interacting domains. The backbone chemical shift assignment determined for the HEAT2 domain in the first

chapter permits us to map the binding sites of both domains of eIF4A onto the HEAT2 domain. We also assign the backbone of the N-terminal domain of eIF4A in order to map on this domain surface, including the HEAT2 domain and a homologous domain from the translational repressor Pdcd4. Finally, we use the binding surfaces of eIF4A-NTD and HEAT2, along with orientational restraints determined using paramagnetic relaxation enhancement, to determine a structure of the two-domain complex. Some of these results have been published in Suzuki et al. 2008 [60] and Marintchev et al. 2009 [42].

Chapter 5 explores larger complexes involving eIF4A. We present work toward an NMR structure of the stable complex formed by full-length eIF4A and the HEAT2 domain of eIF4G. We use single-molecule FRET to measure eIF4A domain separation, in order to understand how different conditions affect the open and closed states of the helicase. We combine many of the pairwise interactions measured in Chapter 4 with other pairwise interactions among domains to construct a larger model of the whole assembly. From this platform, we also describe the mechanism by which eIF4A binding partners modulate eIF4A activity, which appears to involve a dynamic process of subtle alterations in the separation of the two domains. Many of these results have been published in Marintchev et al. 2009 [42].

Finally, in Chapter 6, we discuss conclusions and directions for future investigation.

2

Materials and methods

2.1 Protein expression and purification

2.1.1 Expression constructs

The vector expressing full-length eIF4A, residues 1-406, is a pET-15b derivative containing a hexahistidine tag and thrombin cleavage site [48]. For many experiments in this study, another variant of eIF4A was used, which was less prone to aggregation. This construct, eIF4Ad21, contains an N-terminal HIS tag, followed by a GB1 solubility enhancement tag [72], and a TEV protease cleavage site. It also has a deletion of the first 21 residues of

the wildtype protein, which are not conserved and appear to be unstructured. N-terminal deletion mutants of eIF4A have previously been shown to retain near wildtype activity [38]; In our experience, the GB1-tagged deletion mutant has higher activity than the wildtype, possibly due to the higher solubility.

The eIF4A-NTD is expressed as residues 1-239 in a pETDuet vector with a HIS tag and TEV cleavage site, and was obtained by site-directed mutagenesis to insert as stop codon after residue 239. All constructs containing parts or domains of eIF4G contain an N-terminal GB1 tag, followed by a HIS tag and TEV cleavage site.

2.1.2 Overexpression

Recombinant proteins were expressed in *Escherichia coli* BL21(DE3) in either LB for production of unlabeled protein, or M9 minimal media containing $^{15}NH_4Cl$ and/or deuterated or ^{13}C-labeled glucose. Cells were grown at 37°C to an A_{600} of 0.6-0.8, then transferred to 18-25°C before overnight expression was induced by the addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation and the resulting pellets were stored at -80°C.

Several NMR experiments made use of so-called ILV-labeled samples, which are uniformly $^{15}N^2H$ or $^{14}N^2H$ labeled, with selective protonation and ^{13}C labeling of the methyl carbons of Ile (δ_1 position only), Leu, and Val residues. These were prepared by overexpression in M9 minimal medium in D_2O containing $^{12}C^2H$-glucose. Cells were grown at 37°C until 1 hr prior to induction of overexpression, then transferred to 25°C and supplemented with 125mg/L [3-^{2}H;4,4-^{13}C] α-ketoisovalerate and 75mg/L [3,3-$^2H_2$;4-^{13}C] α-ketobutyrate[20]. α-ketoisovalerate and α-ketobutyrate precursors were purchased (fbreagents.com, Cambridge MA) either in fully protonated form, or already deuterated at the 3-position. The methylene protons of the protonated forms were exchanged to deuterium using a protocol

similar to what has been described previously [19, 20]: each compound was dissolved in 10mL D_2O and titrated to pH 11 by addition of sodium deuteroxide (CIL), and incubated at 45°C for at least 1.5 hr. Reaction progress was monitored by NMR, and the reactions were quenched and returned to neutral pH by the addition of a small volume of 6M HCl. The compounds were stored frozen until ready for use. Figure 2.1 shows the 1D ^{1}H NMR spectra of the α-ketoisovalerate before and after conversion; there is no evidence of significant amounts of aldol condensation products, despite the high concentrations used in the reaction.

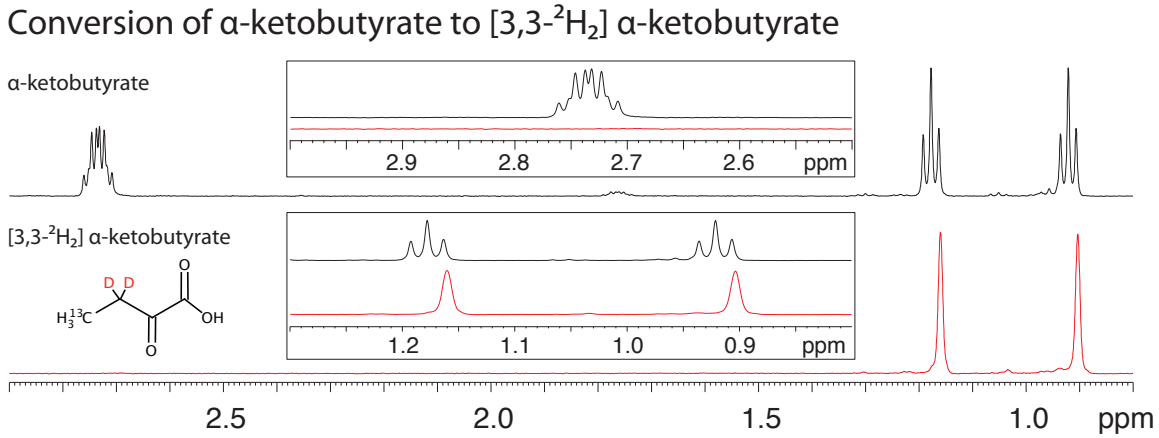

Figure 2.1: 1D ^{1}H NMR spectra of α-ketoisovalerate before (black) and after (red) incorporation of ^{2}H at the methylene position, recorded at 500 MHz. The methylene signal vanishes, and the splitting of the methyl signal is reduced from a doublet of triplets to a simple doublet, caused by coupling to ^{13}C. No aldol condensation products are observed.

Selectively labeled samples of eIF4A-NTD with ^{15}N incorporated into only Ile, Leu, Val, Lys, or Met residues were prepared using a modified minimal medium with individual amino acids and nucleosides replacing NH_4Cl as the nitrogen source. Each 0.5L culture contained 3g Na_2HPO_4, 1.5g KH_2PO_4, 0.25g NaCl, 75mg ^{15}N amino acid, 100mg of each unlabeled amino acid, 63mg adenine, 63 mg guanine, 25 mg thymine, 63mg uracil, 63mg cytosine, 1 mL 1M $MgSO_4$, 0.5 mL 0.1M $CaCl_2$, 0.5mL 0.1% thiamine, 10ml 20% glucose, 0.5mL

50 mg/mL kanamycin, 50μL of a 10000x vitamin mixture, and 20μL of a trace element solution (described in [63]). To minimize scrambling, cells were harvested after only 2 hours of expression.

2.1.3 Purification

Cell pellets were thawed and resuspended in buffer containing 50mM Tris pH 8.0, 200mM NaCl, 10% glycerol, 2mM β-mercaptoethanol, 0.1% triton x100, and 0.4mg/mL lysozyme, and 250U Benzonase nuclease (Novagen). Cells were lysed by incubation with lysozyme at 4°C for 20 min, followed by brief sonication at low power. After high speed centrifugation, the recombinant protein was purified from the supernatant by Ni^{2+} affinity chromatography, optionally followed by TEV cleavage, and then size-exclusion chromatography on either a superdex S75 or S200.

Precipitated eIF4A-NTD was unfolded in buffer containing 8M urea, 20mM Tris pH 8.0, 5mM DTT, and then refolded by rapid dilution into a large volume of refolding buffer containing 20mM Tris pH 8.0, 5mM DTT, and 5% glycerol.

2.2 Acquisition of NMR data

NMR spectra were recorded at 298 K on a Bruker 900, 800, 750, 600, or 500 MHz, or Varian Inova 600 or 500 MHz spectrometers, equipped with cryogenic probes. Samples for NMR measurements generally contained 0.2 2 mM protein in buffer containing 20 mM Tris-HCl (pH 7.0), 150 mM NaCl, 2 mM DTT, 1 mM EDTA, and 5% D_2O. Samples for ^{13}C-edited NOESY spectra were prepared in the same buffer, but with deuterated DTT and EDTA, and 100% H_2O, then were lyophilized and resuspended in 99.99% D_2O. Using protonated tris resulted in a few protein peaks being obscured in the aliphatic ^{13}C-edited NOESY-HSQC, but using the same protonated tris stock for all samples made it easier to achieve

consistent pH and peak positions in all spectra, and in this case it seems that the benefits of the protonated tris outweigh the drawbacks.

The 4D methyl-methyl HMQC-NOESY-HMQC spectrum was collected following an unbiased random schedule, sampling 25% of the Nyquist grid. The nonuniformly-sampled 3D NOESY-HSQC was recorded using a schedule following a Poisson Gap sampling pattern [27].

All spectra were processed using NMRPipe [11]. Nonuniformly sampled NOESY spectra were reconstructed using either Iterative Soft Thresholding [28] or Forward Maximum Entropy reconstruction [26].

The software application CARA [32] was used for analysis of 3D spectra, including backbone assignments of both eIF4G-HEAT2 and eIF4A-NTD, and for maintenance of repositories of NMR spectra and data. 2D spectra used for binding studies were analyzed using either CARA or Sparky [21], and 1D spectra were analyzed using MestRe Nova [1]. NOEs were assigned using a combination of manual assignment in CARA [32] and automated assignment in CYANA [25].

2.3 Structure determination

Initial structure calculations were performed using the torsion angle dynamics software CYANA [24]. The main inputs to these calculations were integrated peaklists from a ^{15}N-dispersed NOESY and a ^{13}C-dispersed NOESY, including both assigned and unassigned NOEs, which were automatically assigned using the macro noeassign, based on an earlier CANDID macro [25].

The automatic NOE assignment process takes into account, but does not alter additional restraints which may be provided. In this case, structure calculations also included 64 hydrogen bond restraints, 255 manually-calibrated upper limit distance restraints from the 4D methyl-methyl NOESY spectrum, and 7 manually-calibrated upper limit distance restraints

from an aromatic ^{13}C-edited NOESY. Dihedral angle restraints for ϕ and ψ angles, derived from TALOS [57], and for $\chi 1$ angles, determined from an HNHB spectrum [3], were also provided.

The noeassign process began with peaklists containing 9099 manually assigned NOE peaks and 1431 unassigned NOE peaks, and culled out peaks which gave rise to redundant distance restraints or which could not be consistently assigned, resulting in a final set of 5592 upper distance restraints, and a heavy atom RMSD of 0.68Å for the 20 lowest energy conformations out of 100 calculated structures.

These automatically assigned distance restraints were then used for further structure refinement using Xplor-NIH software [54, 55], along with the angle restraints, hydrogen-bond restraints, and manually-calibrated distance restraints. Some stereospecific Hβ assignments arising from analysis of the HNHB spectrum were also included.

2.4 Site-directed spin labeling

Site-directed mutagenesis was used to prepare a panel of single-cysteine mutants. Each of the three native cysteines was first mutated either to alanine or serine, based on solvent accessibility, and a single residue on the surface was mutated to cysteine. Some attempt was made to avoid mutating conserved residues, and no mutations were made in sites believed to interact closely with eIF4A-CTD. Each mutant was prepared according to the standard protocol with no isotope labeling, except for exchange into buffer without reducing agent after TEV cleavage. The protein was then concentrated to a volume less than 2.5 mL. A ten-fold excess of the nitroxide spin label (1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl) Methanethiosulfonate (MTSL; Toronto Research Chemicals) was added from a 100mM stock in acetonitrile. The protein was incubated overnight with MTSL at 4°C, then excess spin label was removed by a PD10 desalting column, followed by size-exclusion

chromatography.

The spin-labeled eIF4A-NTD was combined with $^{2}H^{15}N$-labeled eIF4G HEAT2. A ^{15}N TROSY spectrum was recorded on the oxidized form of the complex. The nitroxide spin label was then reduced by the addition of 10mM ascorbic acid, from a 500mM stock in Tris pH 7, with 1 hour incubation at 25°C. A final ^{15}N TROSY spectrum was recorded on the reduced form of the complex.

2.5 Analytical size-exclusion chromatography

Analytical size exclusion chromatography was performed on an analytical Superdex 200 (GE Healthcare) column, with running buffer containing 20 mM Tris (pH 7.0), 150 mM NaCl, 2 mM DTT, 5% glycerol, 5 mM MgCl2, and 0.5 mM EDTA. Except where indicated otherwise, the buffer also included 1 mM ATP. Injected samples contained each indicated protein at a 10 μM concentration, along with 10 ?M FITC-U_{40} and 1 mM AMPPNP. 0.5 ml fractions were collected beginning at 7 minutes after injection, and analyzed by SDS-PAGE.

2.6 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) binding experiments were carried out on a VP-ITC machine (MicroCal) at 25°C by injecting aliquots of eIF4G HEAT2 into a well containing full-length eIF4A with and without HEAT1, in the buffer described above for NMR experiments, except that DTT was replaced with TCEP, and 5 mM $MgCl_2$ and 5% glycerol were also included. Data were processed using Origin software.

3

NMR structure of the eIF4G HEAT2 domain

HEAT domains are common mediators of protein-protein interactions, consisting of three or more copies of the HEAT repeat motif, packed against one another in series. The motif is named for the original domains where it was found: **H**untingtin, **E**longation factor 3, a subunit of protein phosphatase 2**A** (PP2A), and **T**arget of rapamycin. There is little sequence similarity among HEAT repeats, but each consists of a pair of antiparallel α-helices, often 10-20 residues in length. The first helix often contains a kink, and is connected to the second helix by a short linker. Many HEAT domains contain a a large number of

repeats, with sequential pairs of helices stacked at small angles like 20°, leading to a long, rod-like structure. The crystal structure of the C-terminal region of eIF4G containing the HEAT2 and HEAT3 domains shows much more compact structures, with interrepeat angles around 45° [5].

The aim of this chapter is to determine the solution structure of the second HEAT domain of eIF4G using nuclear magnetic resonance spectroscopy (NMR). The primary purpose is educational: this domain is exceptionally stable and well-behaved, making it an excellent model system, both for learning the fundamentals of spectroscopy, and for development of new techniques. At 22 kDa, the domain is considerably larger than some other model samples like ubiquitin or GB1, and yet HEAT2 is small enough that deuteration is rarely necessary. The entirely α-helical secondary structure presents a certain challenge, since backbone contacts contribute very little to the global fold. At this size, complete sidechain assignments of the critical aliphatic and aromatic residues can be obtained, but with some difficulty due to problems of signal overlap and transverse relaxation.

Many protein structures have already been determined by both NMR and x-ray crystallography; these examples help us to better understand the relationship between solution and crystal structures. This includes not only elucidating the relative merits and drawbacks of each approach, but also clarifying how to interact with both types of data simultaneously. Computational projects frequently limit their scope to crystal structures to avoid dealing with the complexities of comparing NMR structural data with crystallographic data. This strategy biases the sample set, since some proteins cannot be crystallized, while others are unsuitable for NMR. Adding to the repertoire of protein structures which have been solved by both methods will help bring us closer to the goal of working with the two data types interchangeably.

Occasionally an NMR structure is significantly different from the crystal structure of

the same protein, as can happen when crystal packing induces domain swapping between neighboring molecules, which would not occur in solution. More often, crystal packing artifacts are more subtle, as might occur when a flexible region at an artificial interface is rigidified. In this case, the construct is somewhat different, since HEAT2 was crystallized along with the third HEAT domain of eIF4G, in a single chain. The linker between the two domains is flexible and not visible in the crystal structure, but the presence of an additional domain could have some impact on the fold. Visual inspection of the structure reveals a probable packing artifact: a surprisingly long N-terminal helix extends into a cavity in a neighboring molecule of HEAT2, as shown in Figure 3.1.

This chapter begins with the backbone assignment of the HEAT2 domain, which will be used not only for the structure determination later in this chapter, but also for the binding studies which are the topic of the next chapter. We then discuss what structural insights can be obtained from the backbone assignment alone, including a preliminary secondary structure assignment and more precise domain boundaries. We use then use this information to optimize the construct, trimming away superfluous residues at the N- and C-termini that do not contribute to the fold, but have strong signals that can overwhelm and obscure resonances that are involved in important long-range contacts. The new construct permits the assignment of most aliphatic sidechain resonances, and also many aromatic sidechains, as detailed in Section 3.3. These sidechain assignments are used to generate interproton distance restraints, which are combined with dihedral angle restraints and hydrogen bond restraints for the structure calculation. Finally, Section 3.6 compares the solution structure with the previously determined crystal structure,

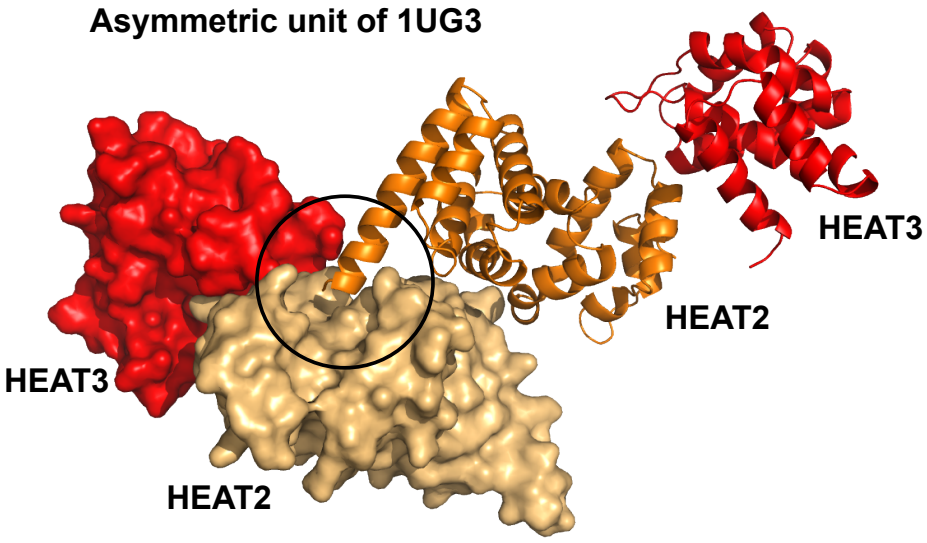

Figure 3.1: Asymmetric unit of the crystal structure of the C-terminal fragment of eIF4G containing the HEAT2 and HEAT3 domains. Both copies of HEAT2 are colored orange, while both copies of HEAT3 are colored red. Chain A is shown as ribbons, and chain B as surfaces. The N-terminus of HEAT2 in chain A appears to form an elongated helix by packing into a pocket formed by HEAT2 and HEAT3 of chain B.

3.1 Backbone assignment of HEAT2

The first step in solving a protein NMR structure is to perform a sequence specific resonance assignment, determining the chemical shifts of the backbone atoms for each amino acid residue in the protein. A complete assignment of all α-, β- and carbonyl carbon chemical shifts, plus non-proline amide proton and nitrogen shifts was obtained using three traditional pairs of 3D triple resonance experiments on a 1mM uniformly $^{2}H^{13}C^{15}N$-labeled sample of eIF4G HEAT2. The three pairs are the HNCA with HN(CO)CA, the HN(CA)CO, with HNCO, and the HN(CA)CB with HN(COCA)CB. These pairs correlate amide proton and nitrogen resonances with Cα, carbonyl, and Cβ chemical shifts, respectively. The first experiment in each pair gives a signal for both the intraresidue carbon (i) and the carbon atom of the previous residue ($i-1$), with the latter typically being weaker. The second ex-

periment gives only the interresidue correlation. Using all three pairs of experiments helps to establish unambiguous sequential connectivity relationships by distinguishing among potential neighbors in cases of chemical shift degeneracy. This becomes increasingly important in assigning larger proteins.

The software program CARA [32] facilitates the sequential assignment in a semi-automated way by two means. It helps to organize peaks from each triple resonance spectrum into unassigned "spin systems", with shared amide proton and nitrogen frequencies. For each spin system, it suggests potential predecessor and successor spin systems, based on matching i and $i-1$ frequencies for each carbon type in a predecessor-successor pair. Strips of the relevant spectra are displayed for visual verification of each suggested match, allowing the user to manually link spin systems to their successors and predecessors.

Once a string of consecutive spin systems has been linked, characteristic Cα and Cβ chemical shifts for particular amino acid types are used to assign those spin systems to a particular stretch of the amino acid sequence of the protein. CARA again provides a degree of automation to this step by computing the probability that each segment of the user-provided protein sequence could have the same chemical shifts as in the string of unassigned spin systems, and suggesting the most probable matches, based on data from the BMRB. Again, the program displays potential matches for visual verification, allowing the user to opt to link more spin systems before making the sequence assignment, if the probabilities of the top choices are too similar.

The backbone assignments are extremely useful, even outside the context of a full structure determination. In the next chapter (Section 4.1), the backbone assignment of the HEAT2 domain is used to determine which residues are involved in interacting with binding partners such as the individual domains of eIF4A. When combined with structures, this information permits us to map binding surfaces, and ultimately model the complex.

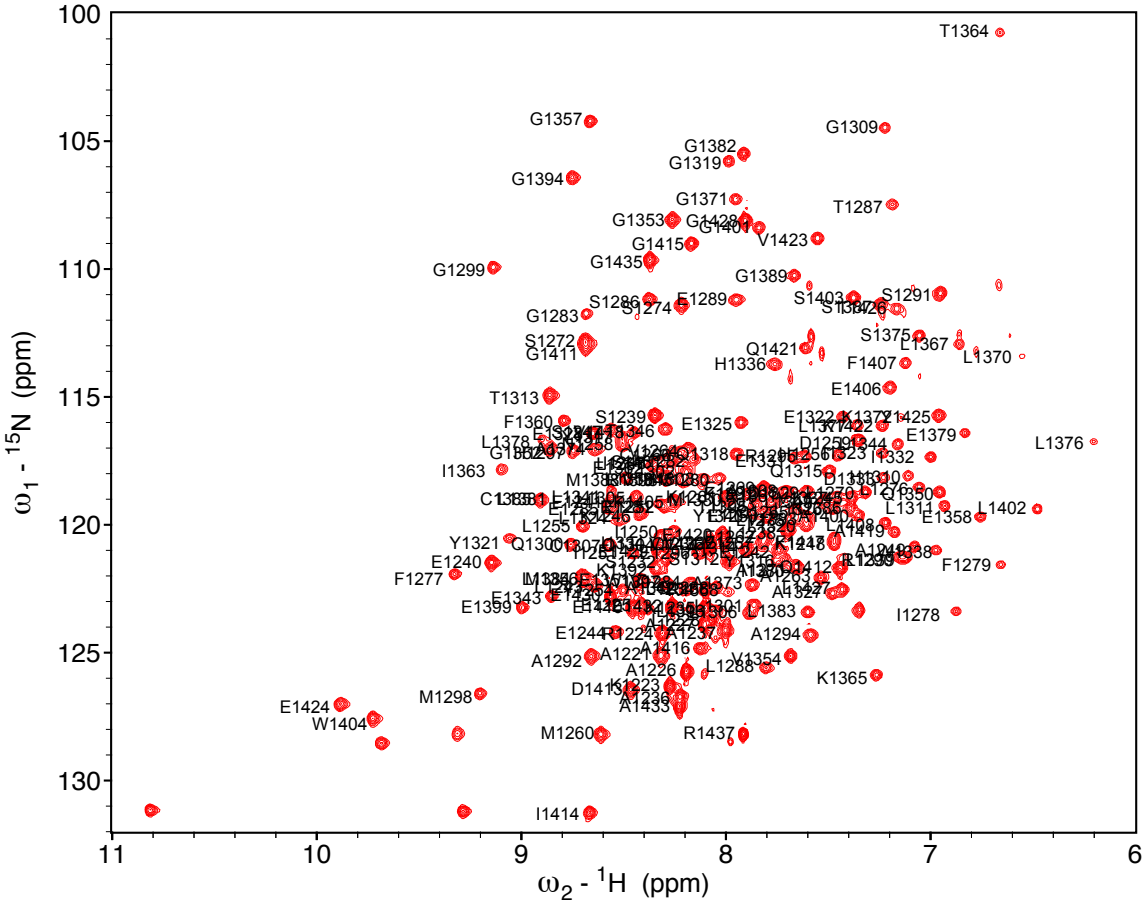

Figure 3.2: ^{15}N-HSQC of eIF4G-HEAT2 with amide backbone assignment. Unlabeled peaks correspond to tryptophan and arginine sidechains.

3.1.1 Nitrogen assignment for prolines

Typically, the backbone assignment is left incomplete for proline residues; the carbon atoms are assigned because they show up as intraresidue correlations in the standard backbone assignment experiments, but the proline nitrogen resonances are not observed because they rely on transferring magnetization from nitrogen to the amide proton for detection, and proline has no amide proton. Some less commonly used backbone assignment experiments, like the HCaN rely instead on Hα detection, and therefore permit assignment of the proline nitrogen atoms. Perhaps more importantly, experiments such as HCaN permit continuous backbone assignments, without the breaks at proline residues and prevent backbone assignment of residues that occur between a pair of prolines.

We used the HEAT2 domain for development of a ^{15}N-detected 2D hCaN experiment that can assist backbone assignment via N and Cα correlations resonances quickly and with extremely high resolution in the nitrogen dimension. This experiment is especially useful for small, unfolded proteins, which have much better signal dispersion in the CαN plan than in the NH plane. Moreover, the ability to continue assignments across proline resides is very helpful for proline-rich peptides, which are often of interest due to their involvement in signaling pathways.

Comparable nitrogen resolution can in principle be achieved in a conventional, proton-detected 3D experiment, but the time required to sample so many points in the nitrogen dimension would be impractical. We showed that nonuniform sampling makes recording the high resolution 3D possible, but the 2D experiment is faster if Hα assignments are not needed [18].

Figure 3.3 shows a 2D hCaN spectrum of HEAT2, which was recorded in 15h 30 min, demonstrating that this pulse sequence is still applicable for proteins as large as 22kDa, although the spectral crowding is a limitation. Lines show a short segment of intra- and

inter-residue connectivity used for sequential assignment, including a connection from a proline residue to its predecessor. The inset shows all the proline assignments for the domain.

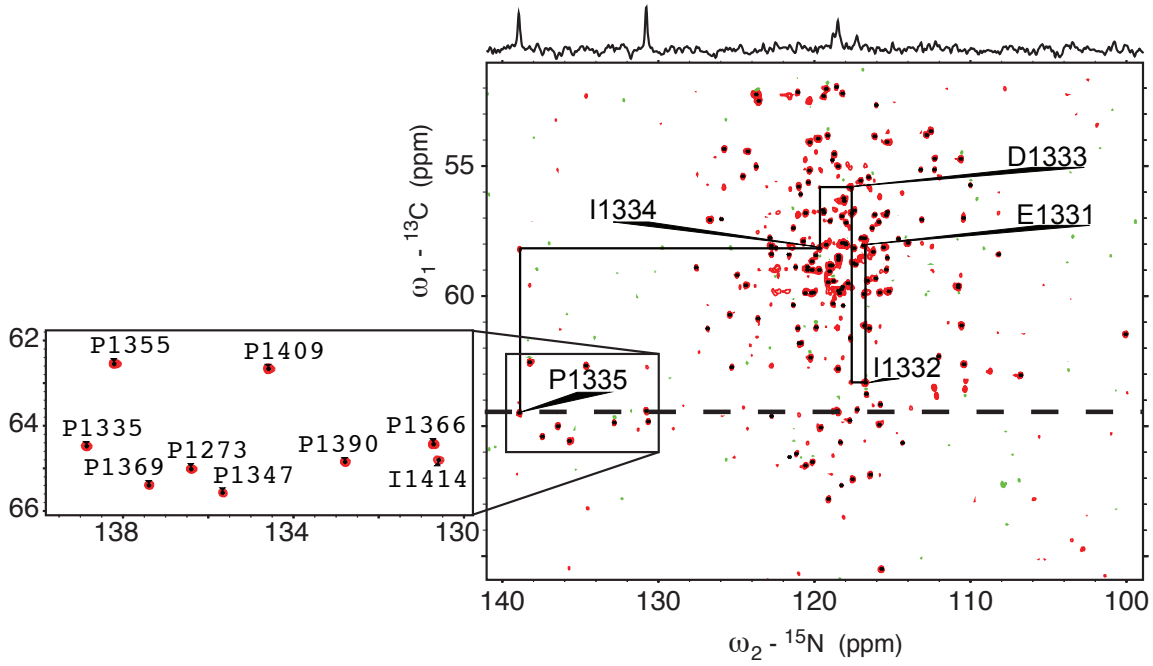

Figure 3.3: 2D hCaN experiment for backbone assignment including prolines. The zoomed region on the left highlights the assigned prolines. Solid lines show a short segment of intra- and inter-residue connectivity used for sequential assignment. The dashed line shows the source of the 1D slice shown along the top, to illustrate sensitivity.

3.2 Secondary structure determination and construct optimization

3.2.1 Secondary chemical shifts

Because chemical shifts are sensitive to local conformation, the backbone assignments of a protein can provide some preliminary indicators of structure. In particular, secondary structure causes consistent upfield or downfield changes in chemical shifts for each backbone atom type and residue type. C^{α} and the carbonyl carbon resonances are shifted downfield in

α-helices, and are shifted upfield in β strands and extended conformations [67]. Meanwhile, H^{α} and C^{β} shifts have the opposite sensitivity to secondary structure. In the first three cases, the changes in chemical shift, also known as secondary chemical shifts, are quite pronounced relative to the width of the overall distribution, so that chemical shifts for helices and sheets cluster separately. On the other hand, C^{β} shifts are less predictive due to overlap in the distributions for each structural motif. Figure 3.4 shows the secondary shift values for Cα and CO for the N-terminal half of HEAT2. As expected for a HEAT domain, these backbone chemical shifts suggest a series of α-helices. The predictions for the exact start- and end-points of the helices are not perfectly reliable, but the first helix appears to start with residue 1240. Before this point, all secondary chemical shifts are small, close to random coil values suggesting that this region is unstructured and flexible.

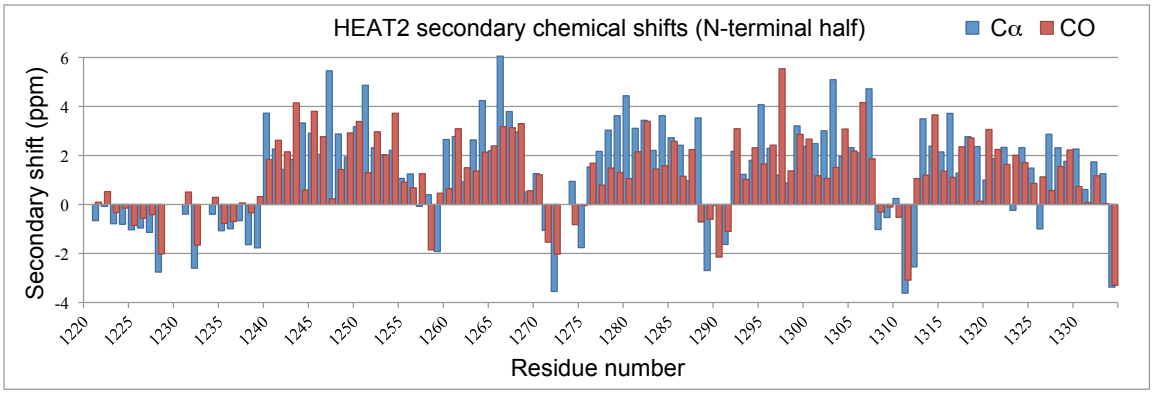

Figure 3.4: Chemical shift deviations from random coil values for Cα (blue) and CO (red) show a series of α-helices, beginning at residue 1240.

These secondary structure predictions can be useful for refining our construct before proceeding with the structure determination. In particular, they help us assess whether unnecessary residues can be trimmed from either terminus of the construct, without loss of important structural elements. These data suggest that the first 19 residues of the original construct can be removed, although the crystal structure shows the helix extending to

residue 1235. The secondary shifts give a less clear prediction for the C-terminus, in part because the last helices in the domain are shorter, and do not form a canonical heat repeat.

3.2.2 Relaxation

Relaxation measurements including T1, T2, and heteronuclear NOE, shown in Figure 3.5 also support trimming the ends of the HEAT2 construct. T2 values average 44ms in the core of the protein, which is surprisingly short for a 22kDa protein, suggesting that there may be some transient oligomerization in the 1.9mM sample. On the other hand, residues at the N- and C-termini have high very T2 values, well over 100ms, as well as low T1 values, and even negative heteronuclear NOE values, all consistent with unstructured peptides tumbling quickly in solution. These parameters are very consistent throughout the core of the domain, suggesting there are no internal regions of exceptional flexibility, but a close inspection of the NOE values shows a decreased NOE for residues around 1335, and again around 1350 and 1410, which are likely to correspond to interhelical loops that are slightly longer and more flexible than the others.

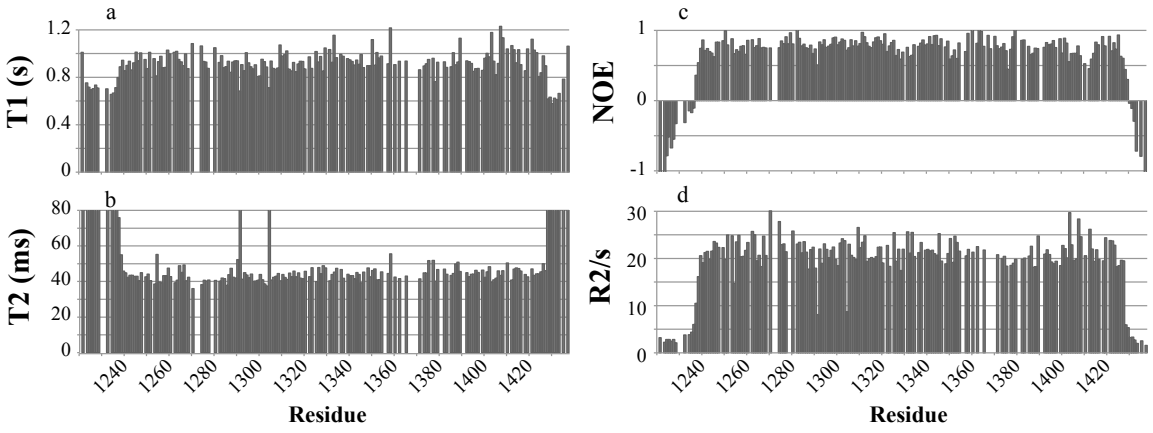

Figure 3.5: Relaxation measurements on HEAT2. Low T1 values (panel a), high T2 values (b), and negative ^{15}N-^{1}H heteronuclear NOEs (c) at the termini suggest flexible tails. High T2 and low NOE values are cut off from the charts. Panel d shows R2 values, which give an indication of the magnitude of the cut off T2 values.

3.2.3 Optimizing the construct

Our goal was to remove the flexible regions flanking the domain, which do not participate in the overall fold of the protein, but whose strong signals would obscure those of folded residues. Site-directed mutagenesis was used to introduce a stop codon after residue 1427, the final residue in the folded domain, in order to remove the subsequent 10-residue linker which would lead to the HEAT3 domain in the full-length protein. Site-directed mutagenesis was also used to make an N-terminal deletion of 16 residues, resulting in a construct beginning with residue A1236, after the glycine residue which remains after cleavage by TEV protease. The N-terminal boundary was made deliberately generous, in order to retain the possibility of forming the extended α-helix seen in the crystal structure. An additional histidine residue was also added to the hexahistidine tag to improve affinity to Ni-NTA during purification.

Figure 3.6 shows some examples of the improvement in spectrum quality caused by deletion of the unstructured N- and C-terminal tails from the construct. The left panel shows the 1D ^{15}N-HSQC of the original HEAT2 (red) and the shorter construct (blue). The latter spectrum has much more consistent signal intensities, without the strong signals between 7.5 and 8.5 ppm resulting from the very mobile residues in the tails. The right panel shows the improvement in signal dispersion in an overlay of the 2D ^{15}N-HSQC of the new construct (blue) over the original HEAT2 (red). Most of the 26 peaks that are lost from the original spectrum come from the crowded central region, as expected for residues in random coil conformations.

The improvements in dynamic range and spectral crowding are even more pronounced in the methyl region of some spectra, such as ^{13}C-dispersed NOESY-HSQC spectra or HCCH-tocsy, or even an ILV ^{13}C-HSQC, where the strong signals in the center of the methyl region from two valine and two leucine residues in the N-terminal tail give rise

to artifacts and completely overwhelm many other signals. Because methyl packing is so critical for defining the fold of an entirely α-helical protein, such artifacts would severely hinder structure determination.

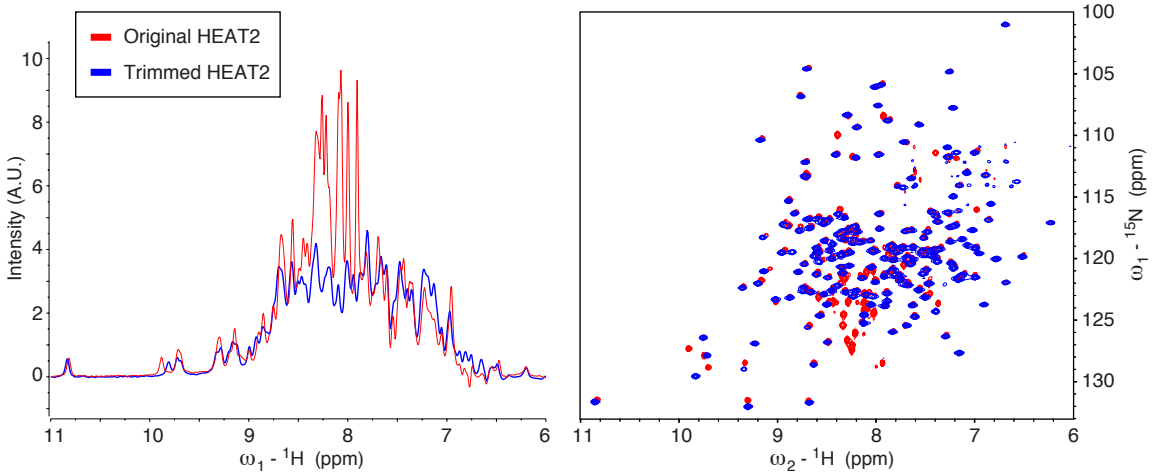

Figure 3.6: The 1D ^{15}N-HSQC spectra (left panel) show that trimming flexible regions from the N- and C-terminus of the construct improves the dynamic range of the spectrum by eliminating residues with exceptionally strong signals. The overlaid 2D ^{15}N HSQC spectra (right panel) show that signal dispersion is also improved. The signals from the unstructured tails mostly fall in the crowded central region of the spectrum.

3.3 Sidechain assignment

Preliminary sidechain assignments for many of the H^{α} and H^{β} atoms were obtained prior to construct optimization using the HBHA(CO)NH experiment, which correlates these protons with the amide proton and nitrogen of the previous residue. Instead of repeating this experiment to verify the assignments in the shorter construct, we opted to use the HNHA and HNHB experiments [3, 64]. These experiments are much less sensitive, so that not every expected crosspeak can be located, but they also give information on the ϕ and χ_1 dihedral angles, respectively. Moreover, both experiments make intraresidue correlations, providing the missing assignments in residues preceding proline. The dihedral angle measurements

are discussed in more detail in Sections 3.4.1 and 3.6.

Long sidechains were assigned using H(CCCO)NH and (H)CC(CO)NH experiments to correlate correlate the amide proton and nitrogen with the sidechain resonances of the previous residue, for proton and carbon respectively. The HCCH-tocsy and HCCH-cosy experiments were used to complete the assignments by correlating the aliphatic protons to their attached carbons.

Many aromatic sidechain assignments were obtained using a pair of 2D experiments, shown in Figure 3.7. The (Hβ)Cβ(CγCδ)Hδ experiment correlates the Cβ chemical shift with the Hδ chemical shift for histidine, tryptophan, phenylalanine, and tyrosine residues. The (Hβ)Cβ(CγCδCϵ)Hϵ experiment gives the same correlations, plus C^{β}-H^{ϵ} for tyrosine and phenylalanine [68]. The sensitivity of these experiments is poor, particularly for the H^{ϵ}. Regardless, chemical shift degeneracy of the β carbons requires validating each of these assignment guesses in NOESY spectra and an aromatic ^{13}C-HSQC.

3.4 Restraints for structure determination

3.4.1 Angle restraints

Backbone dihedral angles were predicted using the program TALOS+ [57] The software uses six kinds of chemical shift assignments from each residue to predict the ϕ and ψ bond angles, including $^{1}H_{\alpha}$,^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, and the amide proton. The program includes a database of tripeptide fragments of 200 proteins for which a high-resolution crystal structure has been solved and for which nearly complete chemical shift assignments are known. The ϕ and ψ angles for the central residue in the tripeptides are determined from the crystal structure, and are correlated against up to 18 chemical shifts for the residue and its immediate neighbors. For each tripeptide of the target protein, TALOS+ searches the database for tripeptides with similar sequence and chemical shifts. If the best matches to

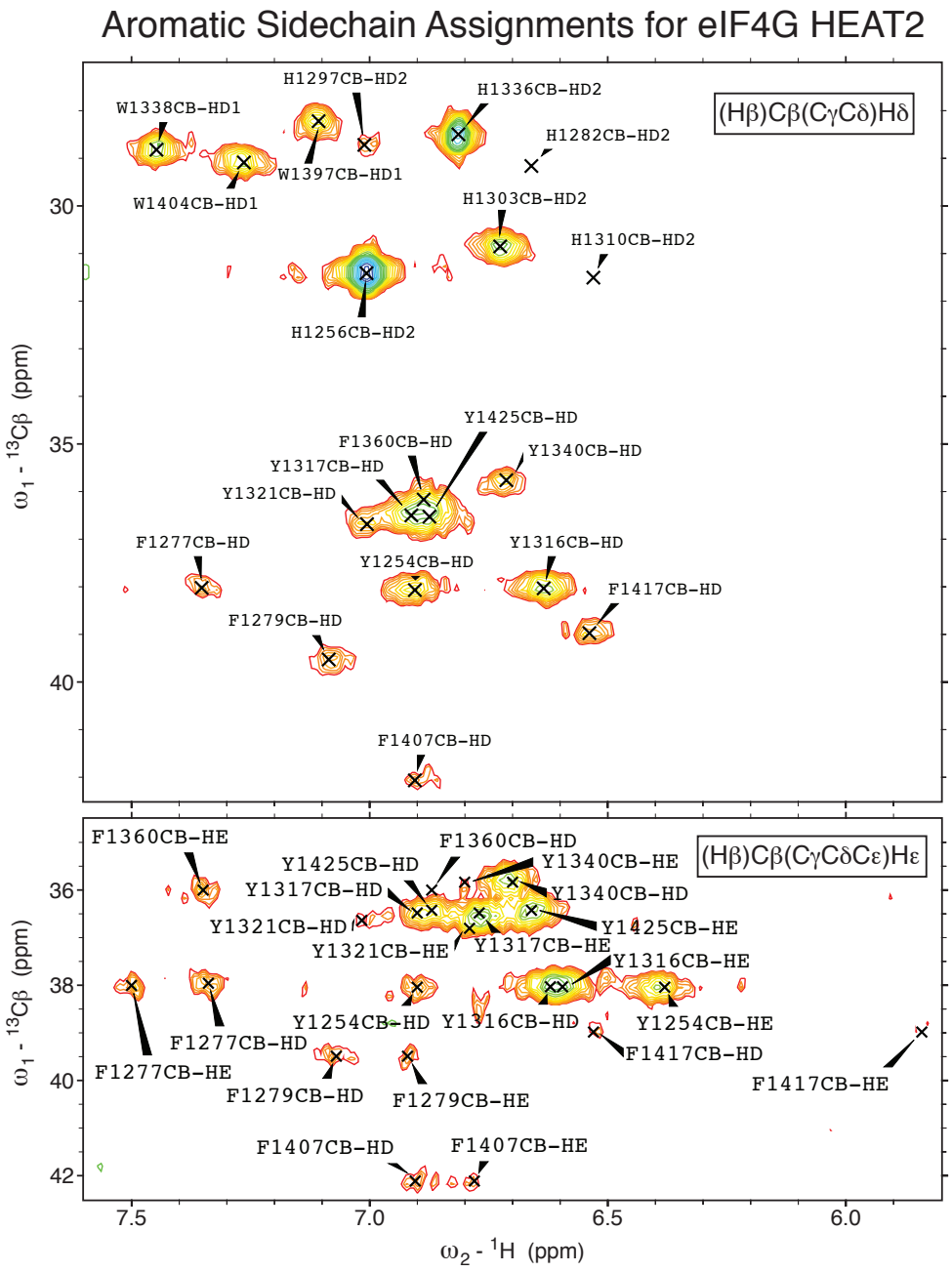

Figure 3.7: Hδ and Hϵ assignments for aromatic sidechains were obtained with the help of two 2D spectra correlating these protons with the corresponding β-carbons. The lower panel shows only the lower portion of the (Hβ)Cβ(CγCδCϵ)Hϵ spectrum because the top halves of the two spectra give the same information.

the chemical shifts have consistent secondary structure or similar backbone angles outside of a secondary structure region of the Ramachandran plot, TALOS+ predicts that the backbone angles for the search peptide will fall within the same range. The ϕ and ψ angles predicted for HEAT2 are shown in the top panel of Figure 3.8, with error bars indicating the variability among the matches in the database.

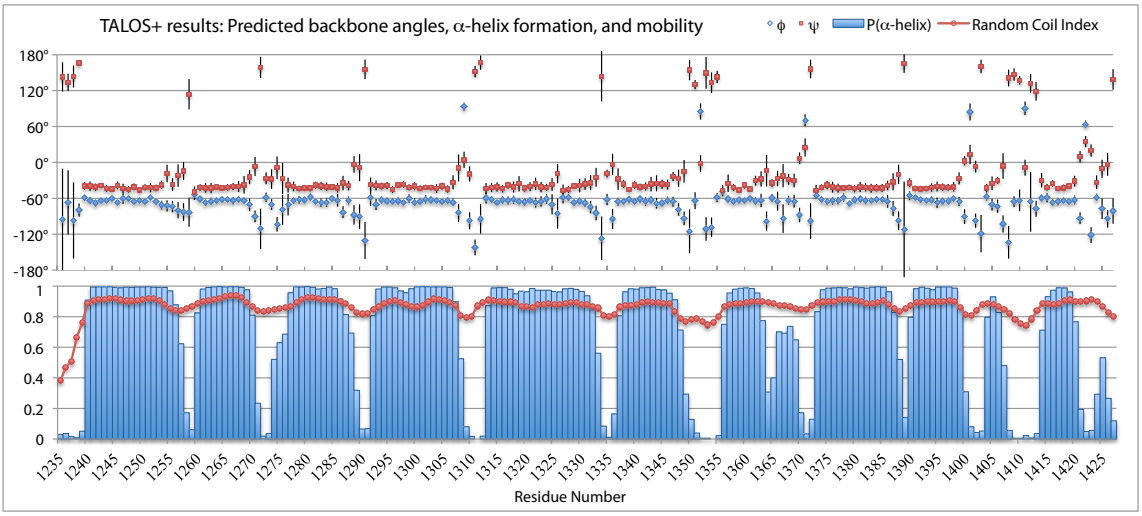

Figure 3.8: The program TALOS+ uses backbone chemical shifts to predict ϕ and ψ angles and regions of secondary structure [57]. In this case, the chemical shifts are consistent with a shorter N-terminal helix than appears in the crystal structure. The blue and red circles in the top panel indicate ϕ and ψ backbone angles, respectively. The blue bars in the bottom panel indicate the predicted probability of α-helix formation for each residue, and the red line in the bottom graph gives the Random Coil Index value, which is an estimate of the order parameter [7].

The bottom chart shows bars indicating the TALOS+ prediction for the probability of α-helix formation at each residue in the sequence. TALOS+ also gives a Random Coil Index (RCI) value for each residue, indicated as a red line on the bottom graph. This is a chemical shift-based estimate of the S^2 order parameter, computed based on the deviation of the provided chemical shifts from random coil reference chemical shifts [7]. This information is of particular relevance because TALOS+ angle predictions should not be used as angle restraints for flexible residues, and it is standard to omit restraints for flexible residues with

an RCI value below 0.5 or 0.65.

In principle, having the assignments for all proline nitrogen atoms could contribute to more reliable angle predictions for residues with neighboring prolines. So far, they are unlikely to be helpful, due to the rarity of proline nitrogen assignments in the existing database. As published, the TALOS+ database contains 23,257 tripeptide fragments, of which 1,177 contain at least one proline residue. However, the nitrogen chemical shift is available in only 10 of those fragments, involving only six different proline residues. On the other hand, the TALOS+ database is designed to be extensible. Adding the bond angles from the crystal structure along with the chemical shifts we have determined for HEAT2 could more than double the number of proline nitrogen assignments available for making predictions for future structural work.

Some χ_1 angle restraints were obtained from a HNHB experiment [3], which gives qualitative information on N-H$^\beta$ J couplings. Comparing the relative intensities of the H$^\beta$ peaks, along with the relative strengths of the H$^\alpha$-H$^\beta$ NOEs and the H^N-H$^\beta$ NOEs can give not only the χ_1 rotamer, but also the stereospecific H$^\beta$ assignment. In total, 44 χ_1 angle restraints and 35 stereospecific Hβ assignments were used in the final structure calculations.

3.4.2 Distance restraints

^{1}H-^{1}H distance restraints were obtained from several NOESY spectra. A 3D ^{15}N-NOESY-HSQC was recorded on a 2mM sample of ^{15}N^{13}C-HEAT2 in H_2O, giving amide-amide and amide-sidechain distance restraints. These amide-amide distance restraints are helpful for defining secondary structure, but in an all α-helical structure such as HEAT2, it shows very few long-range interactions, and therefore is insufficient for determining a global fold.

Most long-range distance restraints were obtained from sidechain-sidechain interactions observed in a pair of 3D ^{13}C-NOESY-HSQC spectra recorded on a 2mM sample of ^{15}N^{13}C-

HEAT2 in D_2O, covering the aliphatic and aromatic regions, respectively. The aliphatic spectrum was recorded in only 18 hours at 800 MHz by using nonuniform sampling.

In order to obtain better resolution and to unambiguously assign NOE crosspeaks between hydrophobic sidechains in the core, an ILV-labeled sample was prepared, fully deuterated and ^{12}C-labeled except for the terminal methyl groups of all isoleucine, leucine, and valine residues [20]. This sample was dissolved in D_2O so that only NOEs between these methyl groups would be observed, and a nonuniformly sampled 4D ^{13}C-HMQC-NOESY-HMQC was recorded. NOEs were assigned using a combination of manual assignment in CARA [32] and automated assignment in CYANA [25].

3.4.3 Hydrogen bond restraints

Hydrogen bond restraints for the final structure calculation were determined using a combination of the chemical shift-based TALOS secondary structure predictions described earlier and deuterium exchange analysis. A 2mM sample of HEAT2 was lyophilized and resuspended in D_2O. A ^{15}N HSQC was immediately recorded, within 30 minutes of addition of D_2O. This spectrum is shown in red in Figure 3.9, overlaid on a spectrum recorded with identical parameters in H_2O. Signals for many residues vanish within the first 30 minutes of contact with D_2O; 100 observable resonances remain, including two peaks from tryptophan sidechains and 98 backbone amide peaks. These observable resonances must belong to residues in which the amide proton is structurally protected from exchange with the surrounding solvent, suggesting the possibility of a hydrogen bond. Hydrogen bond restraints were used between protected amide protons and the carbonyl oxygen at the $i-4$ position in the cases when both residues are predicted by TALOS to occur within an α-helix, and where such restraints were not inconsistent with NOE data.

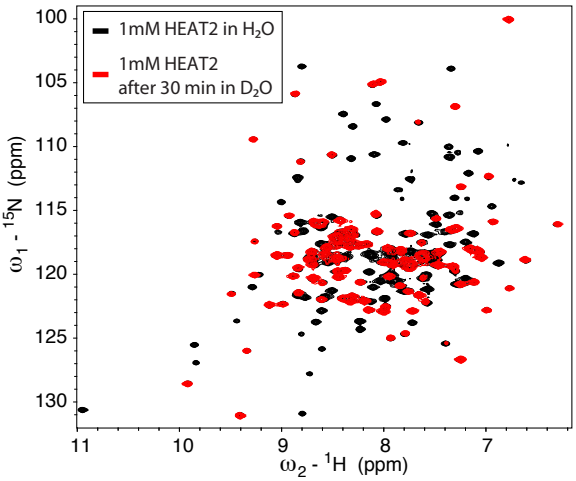

Figure 3.9: Many amide resonances remain in the ^{15}N HSQC of eIF4G HEAT2 after the sample is exchanged into D_2O, indicating which residues are likely to be involved in hydrogen bonds. These data were used in combination with TALOS secondary structure predictions to generate hydrogen bond restraints for structure refinement.

3.5 NMR structure of the HEAT2 domain of eIF4G

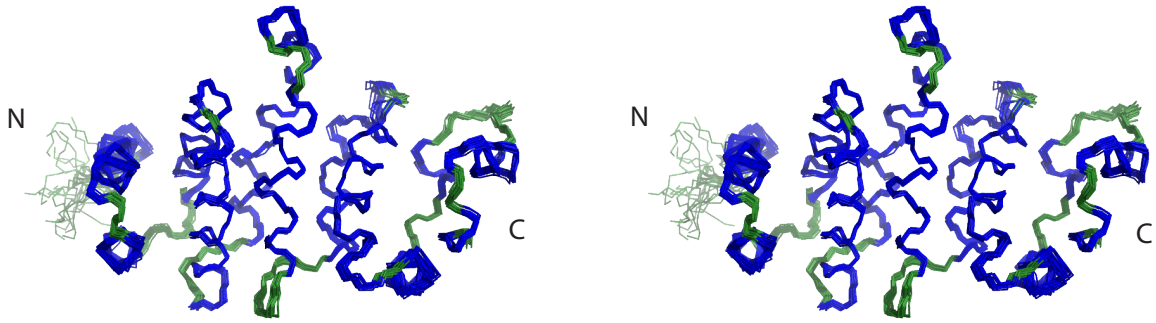

Figure 3.10: Stereoview (crosseye) of the 20 lowest-energy conformers of HEAT2, with α-helices in blue and loops in green.

The solution structure of the HEAT2 domain of human eIF4G is shown in Figure 3.10, as a stereoview of an overlay of the 20 lowest-energy structures from the calculation. Four pairs of long helices are well-defined, capped at the C-terminus by four short helices which might be considered to form a fifth HEAT repeat. Most of the core of the protein is formed by knobs-

into-holes style packing of hydrophobic sidechains such as leucine, isoleucine, and valine, except in the final HEAT repeat, where aromatic sidechains dominate the hydrophobic core. The inter-helical loops are also well-structured; the only exceptions are the flexible residues at the N-terminus before the first helix. Excluding these residues, the backbone r.m.s.d. over residues 1240-1427 is 0.32 ± 0.05Å. Table 3.1 lists the statistics of the structure calculations.

NMR distance and dihedral angle restraints	
Distance constraints	
Total NOE	5900
Intra-residue	1254
Inter-residue	4646
Sequential ($\|i-j\|=1$)	1301
Medium-range ($1<\|i-j\|<5$)	1660
Long-range ($\|i-j\|\geq 5$)	1685
Hydrogen bonds	64
Total dihedral angle restraints	364
ϕ	160
ψ	160
χ_1	44
Structure statistics	
Violations	
Max. distance constraint violation (Å)	0.49
Max. dihedral angle violation (°)	4.75
Deviations from idealized geometry	
Bond lengths (Å)	0.012
Bond angles (°)	1.6
Average pairwise r.m.s. deviation (Å)	
Heavy atoms	0.68 ± 0.04
Backbone atoms	0.32 ± 0.05
Ramachandran Plot Summary from PROCHECK	
Residues in most favored regions	3068 (90.8%)
Residues in allowed regions	300 (8.9%)
Residues in generously allowed regions	8 (0.2%)
Residues in disallowed regions	4 (0.1%)

Table 3.1: NMR and refinement statistics for the eIF4G HEAT2 structure, and Ramachandran Plot summary from PROCHECK [36]

Procheck was used to evaluate the quality of the final structure bundle [36]. The results of the torsion angle analysis are summarized in the bottom panel of Table 3.1. 90.8% of the residues have ϕ and ψ backbone angles that fall in the most favored regions of the Ramachandran Plot. 0.1% of residues fall in disallowed regions, meaning that four of the twenty structures in the bundle contain a single residue with disfavored dihedral bond angles. However, the residues that appear in the disallowed regions of the Ramachandran Plot are exclusively in the unstructured N-terminal region of the domain.

3.6 Comparison of NMR structure with crystal structure

As expected, the NMR structure closely resembles the crystal structure, with a nearly identical arrangement of stacked helical hairpins. The only striking difference is the unstructured N-terminus, where the crystal structure has an extension of the first α-helix, likely induced by packing against another copy of HEAT2. Excluding these residues, the average backbone RMSD of the final 20-structure bundle from the crystal structure is 1.54Å, with a heavy atom RMSD of 2.06Å. Figure 3.11 shows an overlay of the crystal structure of HEAT2 with the 20 conformers of the NMR structure.

The backbone chemical shifts, amide-amide NOE patterns, T1 and T2 relaxation rates, and amide heteronuclear NOEs of residues 1236-1239 are all consistent with the flexible conformations shown in the NMR structure, and not with the alpha helix shown in the X-ray crystal structure.

We have also measured the H^N-H^α J couplings, which provide information on the backbone angle ϕ, in an independent way from the chemical shift-based TALOS+ predictions [64], to confirm that the N-terminus of this construct is indeed more flexible than indicated in the x-ray crystal structure. J-coupling values between the amide proton and the alpha proton are computed from the ratio of the integrals of the amide proton diagonal

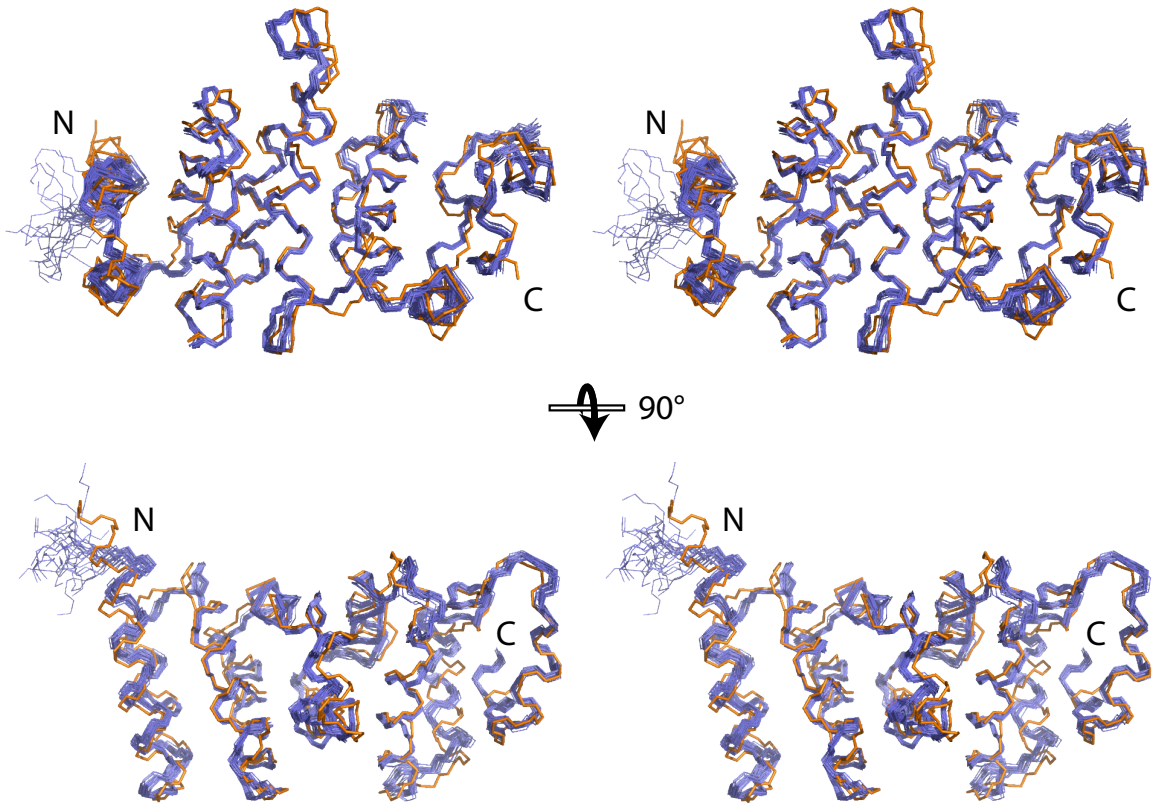

Figure 3.11: Comparison of the HEAT2 NMR structure (blue) with the crystal structure (red).

peak and the alpha proton crosspeak in the HNHA spectrum, using the following equation, where 2ζ is the duration of the coupling evolution:

$$\frac{S_{cross}}{S_{diag}} = -\tan^2(2\pi J_{HH}\zeta)$$

Figure 3.12 shows the couplings measured in the HNHA experiment. ϕ angles consistent with these couplings can be computed using a Karplus equation, $J = A\cos^2(\phi - 60) + B\cos(\phi - 60) + C$, with empirically determined A, B, and C parameters [31, 65]. Small couplings around 4-5 Hz are consistent with the ϕ angles around -60° found in α helices. These results show that the N-terminal α helix begins at residue 1240, not residue 1235 as in the crystal structure.

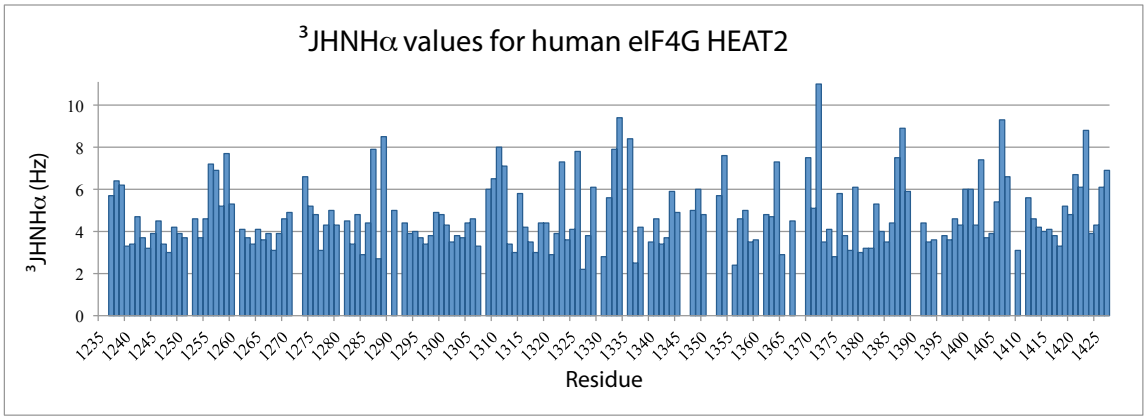

Figure 3.12: ϕ angles can be determined from 3-bond scalar couplings measured in the HNHA spectrum. These data also show that the N-terminal α helix begins at residue 1240, not residue 1235 as in the crystal structure.

The angle restraints generated from the J-coupling values are mostly redundant with those from TALOS+, and were therefore omitted from the structure calculation, but were only used as an independent verification of the helix boundaries.

4

The eIF4G-HEAT2:eIF4A-NTD complex

The eukaryotic initiation factor 4A (eIF4A) binds to the HEAT2 domain of eIF4G primarily through the N-terminal domain, which also contains the catalytic site for ATP hydrolysis and has many of the motifs responsible for binding ATP and RNA. In this chapter, we explore these binding interfaces using NMR. We measure chemical shift perturbations in HEAT2 and in eIF4A-NTD to determine binding interfaces on each molecule. HEAT2, RNA, and the MA3c domain of the translational suppressor Pdcd4 have closely overlapping binding surfaces on eIF4A-NTD, leading us to the conclusion that they compete with one

another for binding. Finally, we construct a model of the HEAT2:eIF4A-NTD complex using the measured binding interfaces along with site-directed spin labeling and paramagnetic relaxation enhancement to determine the relative orientations of the domains.

4.1 eIF4A-NTD and eIF4A-CTD binding interfaces on HEAT2

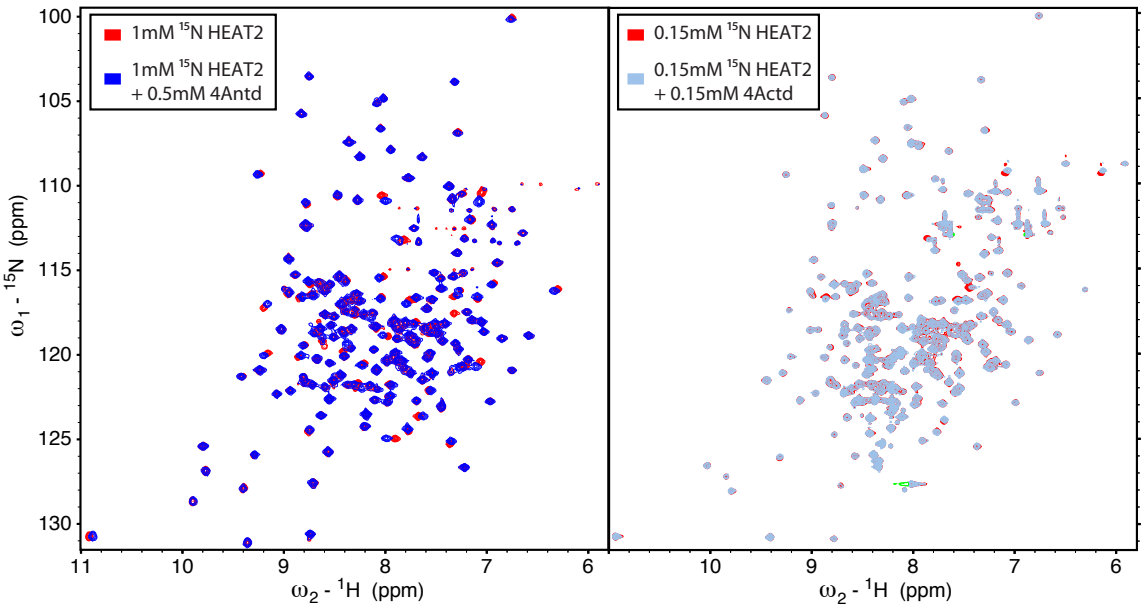

Figure 4.1: ^{15}N-HSQC titration of ^{15}N-HEAT2 with unlabeled eIF4A-NTD (left panel) and with eIF4A-CTD (right panel). In both panels, the spectrum of the free HEAT2 is shown in red. The spectrum of HEAT2 bound to eIF4A-NTD is shown in dark blue, and the spectrum of HEAT2 bound to eIF4A-CTD is shown in light blue.

Chemical shift perturbations show that each individual domain of eIF4A binds to the HEAT2 domain of eIF4G. In each case, ^{15}N-labeled HEAT2 was titrated with the unlabeled domain of eIF4A, and ^{15}N HSQCs were recorded for each step. In the case of eIF4A-CTD (Figure 4.1, right panel), a few peaks in the HSQC of the HEAT2 domain shift only by very small amounts, reflecting a weak interaction with a small interface. In contrast, the spectrum of HEAT2 changes markedly in the presence of eIF4A-NTD (Figure 4.1,

left panel). Many peaks shift, including some that shift substantially, and also weaken significantly in intensity. This suggests a stronger binding interaction, though still weak enough to be in the fast-exchange regime on the NMR time scale.

Using the sequential backbone assignment from the previous chapter, we were able to determine the binding interfaces on eIF4G HEAT2 for both domains of eIF4A. Changes in chemical shifts were measured for each peak that could be tracked, along the ^{1}H and ^{15}N dimensions. The normalized distance in ppm ($\Delta\delta$) was computed as $\sqrt{(\Delta H)^2 + 0.2(\Delta N)^2}$. The normalization factor of 0.2 roughly compensates for the broader range of nitrogen chemical shifts than proton chemical shifts in typical protein ^{15}N HSQCs. The amide nitrogen range is often around 30ppm, while amide protons may span a range of 6ppm. These values are plotted in Figure 4.2. In each titration, residues with shifts greater than one standard deviation above the mean were selected as likely to be involved in binding.

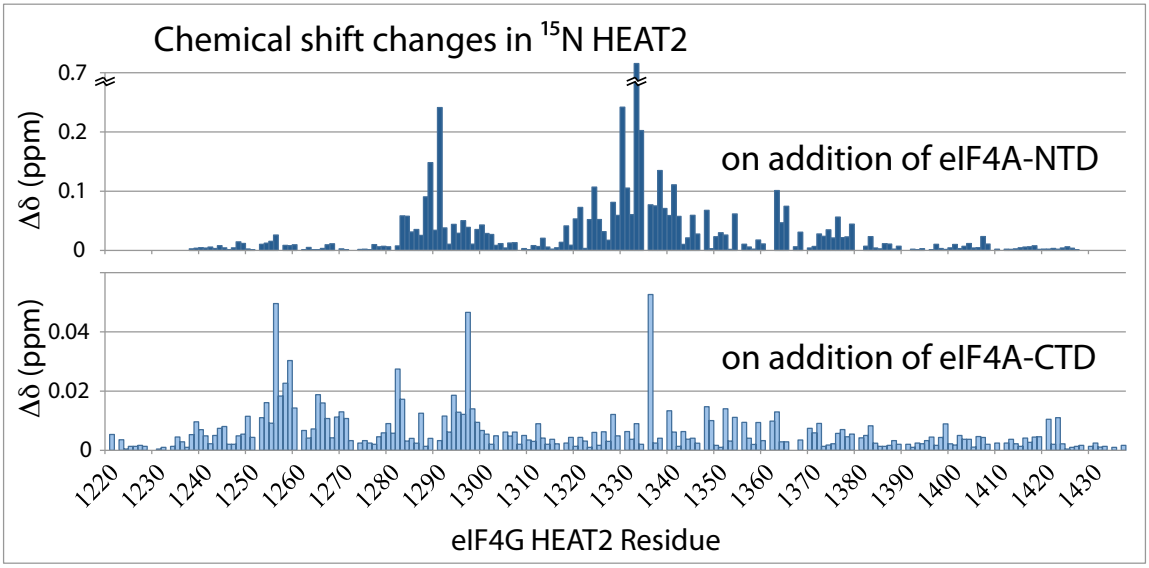

Figure 4.2: Normalized chemical shift changes caused by titration of unlabeled eIF4A-NTD (above) or eIF4A-CTD (below) into ^{15}N-HEAT2.

The chemical shift perturbations caused by the titration of ^{15}N HEAT2 with eIF4A-NTD

and eIF4A-CTD cluster into two regions, as shown in the top panel of Figure 4.3. eIF4A-CTD binds primarily to the loop between the two helices in the first HEAT repeat, with significant effects also observed on the second HEAT repeat. eIF4A-NTD affects the loop of the third HEAT repeat most, and neighboring HEAT repeats also feel some effects of binding. Key residues in the interaction with the CTD are painted light blue, and those that are most affected in the interaction with the NTD are painted dark blue.

The bottom panel of Figure 4.3 shows that these two binding sites correspond to negatively charged regions on the domain surface. Indeed, some of the most affected residues in this titration include D1259, E1289, E1328, and D1333. We can expect to find complementary positively charged surfaces on both domains of eIF4A where HEAT2 binds.

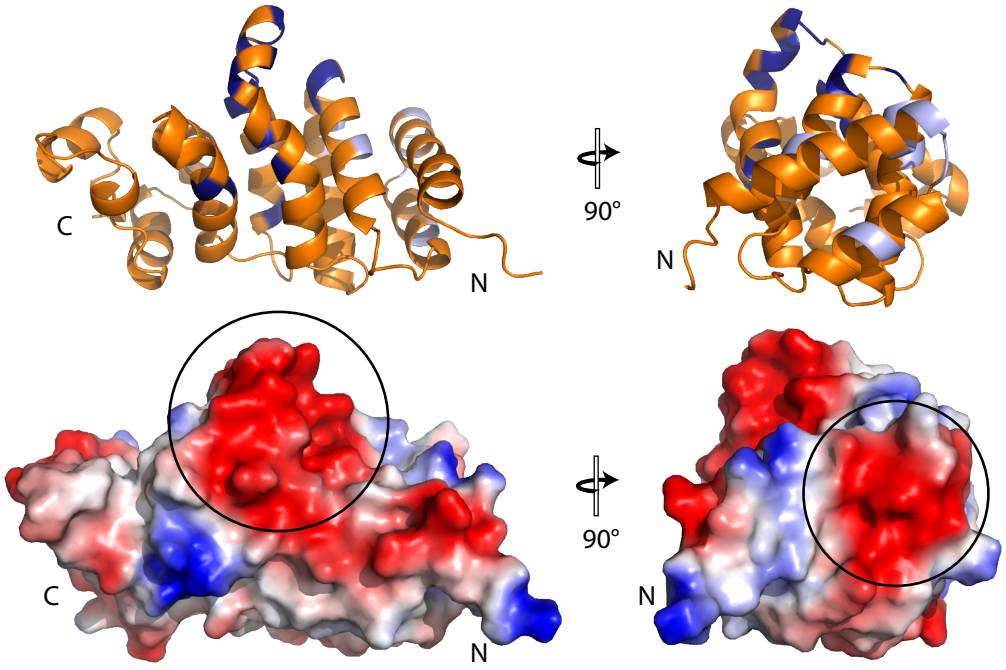

Figure 4.3: eIF4A domain binding surfaces on eIF4G HEAT2. The top panel shows residues with strong chemical shift perturbations upon titration with eIF4A-NTD, painted dark blue, and residues implicated in binding eIF4A-CTD, painted light blue. The bottom panel shows the HEAT2 domain from the same perspectives, as a surface colored by electrostatic potential. Two negative patches are circled, corresponding to the eIF4A-NTD binding surface in the left panel, and to the eIF4A-CTD binding surface in the right panel.

The chemical shift changes caused by eIF4A-CTD are quite subtle, and appear to cluster near histidine 1256, raising some concern that the shifts we measure might be caused by pH changes, rather than by binding. We therefore used site-directed mutagenesis and surface plasmon resonance to verify that the identified region of the HEAT2 domain is involved in binding to eIF4A-CTD. D1259 was identified as part of the eIF4A-CTD interface of HEAT2, and was mutated to asparagine, to assess its influence on the binding interaction. This mutation decreases binding of HEAT2 to full-length eIF4A approximately fourfold, yet has no effect on binding between HEAT2 and eIF4A-NTD. Therefore, we conclude that D1259 is indeed in the eIF4A-CTD binding site of the HEAT2 domain.

It is worth noting that the HEAT2 residues most affected by eIF4A-NTD binding correspond very closely with the residues in Pdcd4-MA3m and Pdcd4-MA3c that are affected by eIF4A-NTD: both of these domains also show strong chemical shift changes in the interhelical loops of HEAT repeats 2 and 3 when titrated with eIF4A-NTD [60]. We were unable to measure eIF4A-CTD binding to either domain of Pdcd4; apparently binding is even weaker than with HEAT2. However, it has been shown that mutation of residue D1259 in HEAT2, which our measurements show to interact with eIF4A-CTD, has a similar effect on eIF4A binding as mutation of the corresponding residues in each MA3 domain of Pdcd4 [70]. Subsequent crystal structures of the eIF4A-Pdcd4 complex confirm that eIF4A-CTD does indeed interact with this site on the interhelical loop of the first HEAT repeat in each domain of Pdcd4 [9].

Although the differences in how Pdcd4 and HEAT2 interact with eIF4A are ultimately more interesting than the similarities, we can in the meantime use the Pdcd4 structures to help our understanding of how HEAT2 interacts with eIF4A. We have used the structure of human Pdcd4 bound to eIF4A [9] to construct a model of how HEAT2 interacts with eIF4A, by aligning HEAT2 with the MA3m domain of Pdcd4. The left panel of Figure 4.4

shows this model. The right panel shows eIF4A only, from the same perspective, so that we can see the basic patch where HEAT2 is expected to bind, matching the acidic patches we see on HEAT2 in Figure 4.3.

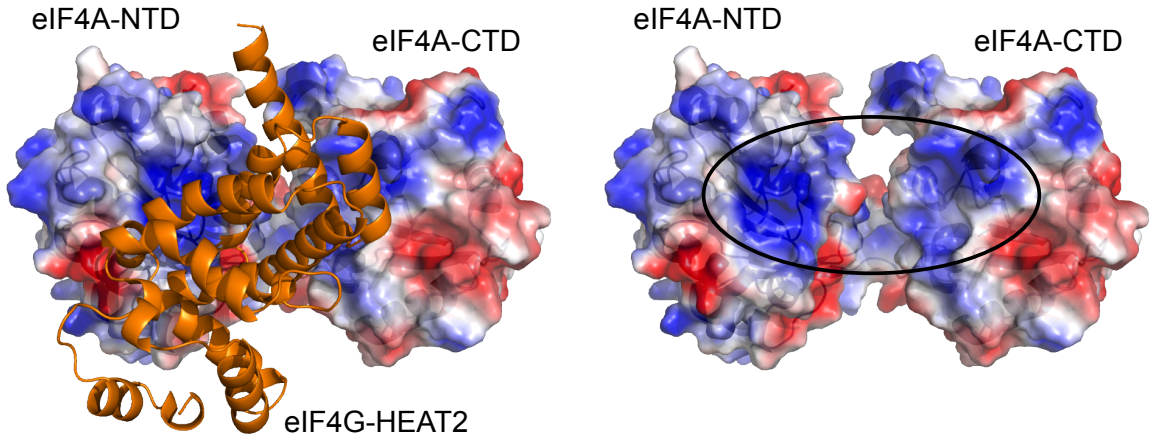

Figure 4.4: Model of HEAT2 binding to eIF4A, based on eIF4A binding to Pdcd4-MA3m (part of PDB 2ZU6 [9]. HEAT2 (from 1UG3 [5] is shown as an orange ribbon, occupying the position of MA3m. eIF4A is shown as a surface, painted by electrostatics. In the right panel, with HEAT2 removed, a basic patch (circled) is evident where HEAT2 is predicted to bind on each domain of eIF4A, complementary to the acidic patches on HEAT2 shown above.

4.2 Backbone assignment of eIF4A-NTD

The backbone assignment of eIF4A-NTD enables us to observe the interaction of eIF4A-NTD with HEAT2 from the opposite perspective, using the same chemical shift perturbation methods, but this time with ^{15}N-labeled eIF4A-NTD and unlabeled HEAT2. At 26 kDa, the N-terminal domain of eIF4A is slightly larger than the HEAT2 domain, making assigning the backbone chemical shifts more difficult. Lower solubility and line-broadening in flexible loops add to the challenge. Fortunately, the crystal structure has been solved (PDB: 2G9N), so we can observe how the interacting residues cluster on the surface of the 3D structure without also solving the NMR structure.

4.2.1 Sample development

Although the individual domains of eIF4A are considerably more soluble than the full-length protein, they nonetheless present significant challenges for obtaining a suitable NMR sample. Our goal was to find the right balance of minimizing sample aggregation during the NMR experiment while maximizing the signal-to-noise ratio and spectraum quality. We have explored a variety of buffer conditions for both purification and measurement.

In many cases, minimizing aggregation and maximizing signal-to-noise are at odds with one another. Decreasing sample concentration, increasing salt concentration, and using relatively high pH all help to minimize aggregation, especially during the purification process, but each strategy decreases the S/N of the NMR spectra. At a moderate salt concentration of 150mM, protein concentrations as high as 0.9 mM can be reached with little aggregation in the short term, but for multi-day experiments at room temperature, 0.5 mM is more suitable, still with some precipitation during the experiment.

The choice to avoid phosphate buffer was more clear; comparing a ^{15}N HSQC of eIF4A-NTD in tris buffer with one in phosphate buffer shows that many peaks that are present in the tris buffer are broadened or even missing in phosphate. In retrospect, this effect is unsurprising, since many residues in eIF4A must interact with the phosphate moieties in RNA and ATP in order for it to perform its function as an ATP-dependent RNA helicase, and these transient interactions cause line-broadening. The most broadened residues were later found to be in the phosphate binding loop, also known as the Walker A motif (such as residues 79-83). Other residues which are broadened and shifted in the presence of phosphate have since been shown to be involved in RNA binding. Several examples of broadened peaks are labeled in Figure 4.5.

At 26 kDa, the N-terminal domain is large enough to require deuteration for the standard backbone assignment experiments. However, the domains of eIF4A are each structurally

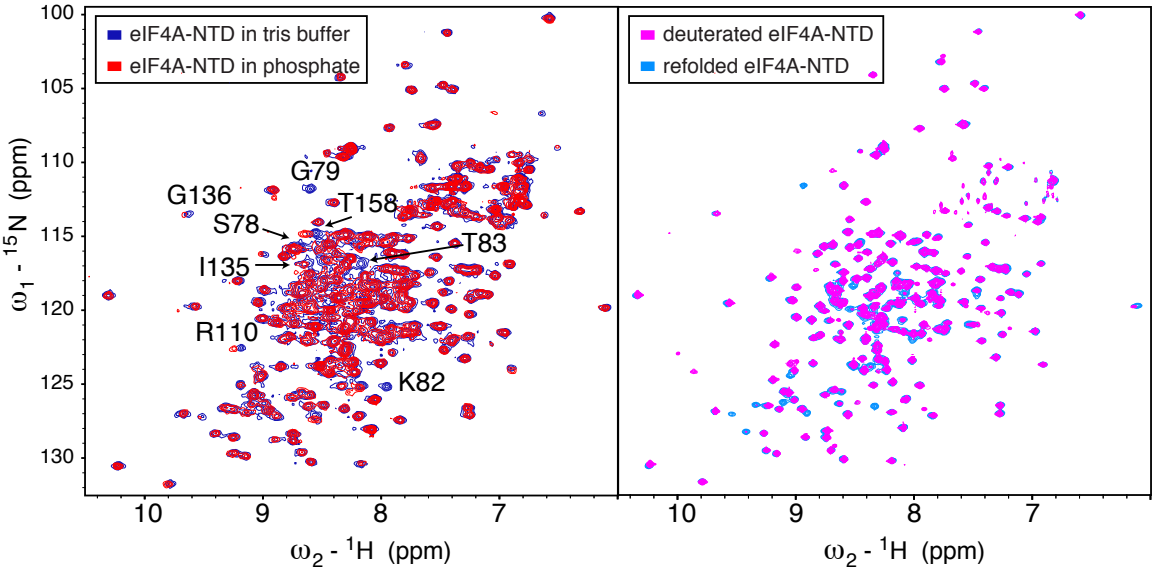

Figure 4.5: Determining conditions for eIF4A-NTD backbone assignment. The left panel shows that many resonances which are present in tris buffer (blue) are broadened or absent in phosphate buffer (red). These resonances were later verified to be involved in ATP or RNA binding; several examples are labeled. The right panel shows the benefits of refolding deuterated eIF4A-NTD. Many peaks which have since been assigned to the interior of the domain are missing from the spectrum before refolding (magenta), but appear after refolding (cyan).

composed of a small central β-sheet, surrounded on both sides by α-helices, and the amide protons in the β-sheet are involved in strong hydrogen bonds that are buried in the core of the protein, completely inaccessible to the solvent. Therefore, many expected signals are missing from the protein after expression in deuterated media. In retrospect, making the optimal sample of eIF4A-NTD for backbone assignments would have included unfolding the protein in a denaturing buffer to allow ^1H_2O to access these residues, followed by refolding. Figure 4.5 shows the signals which could have been recovered by refolding.

4.2.2 Triple resonance experiments

The effort to assign the backbone resonances of the N-terminal domain of eIF4A initially followed the same procedure as outlined above for the HEAT2 domain of eIF4G (Section 3.1).

The HNCA, HN(CO)CA, HNCO, HN(CA)CO, and HN(CA)CB spectra were recorded on a pair of $^{2}H^{15}N^{13}C$-labeled samples. Unfortunately these spectra were insufficient for a complete backbone assignment, mainly for two reasons. First, the amide protons in the central β-sheet of the domain structure have very little solvent exposure due to the surrounding α-helices, and they are involved in strong hydrogen bonds, so that they remain almost exclusively deuterated after being expressed in $^{2}H_2O$ M9 media, rather than exchanging with $^{1}H_2O$ during purification. Therefore, the signals from these residues are extremely weak in all H^N-detected spectra. Secondly, carbon chemical shift degeneracy also contributed to assignment difficulties.

Several additional spectra help to fill in some of the remaining gaps. A CC(CO)NH spectrum was also recorded on a partially protonated sample, providing the missing sequential Cβ resonances, as well as hints for the amino acid type. A ^{15}N NOESY-HSQC spectrum was useful for verifying that an assignment candidate has the expected neighbors in an α-helix, despite some uncertainty in Cα, Cβ, or carbonyl connectivity. Used in combination with the existing structure (2G9N), it also helped to confirm that residues in the central β-sheet have the expected NOEs to residues in a neighboring strand. Finally, amino-acid selective labeling was used to narrow the search space for assignments.

4.2.3 Amino-acid specific labeling

Several fully protonated samples were prepared with ^{15}N incorporated only at certain amino acid types, so that ^{15}N HSQCs could be recorded with only those peaks that correspond to that particular amino acid. Preparation of these samples requires providing a single ^{15}N-labeled amino acid to unlabeled minimal media, and some precautions must be taken to prevent scrambling of the label to other amino acids by transaminases or other metabolic processes. In some cases, auxotrophs must be used. However, for several amino acids,

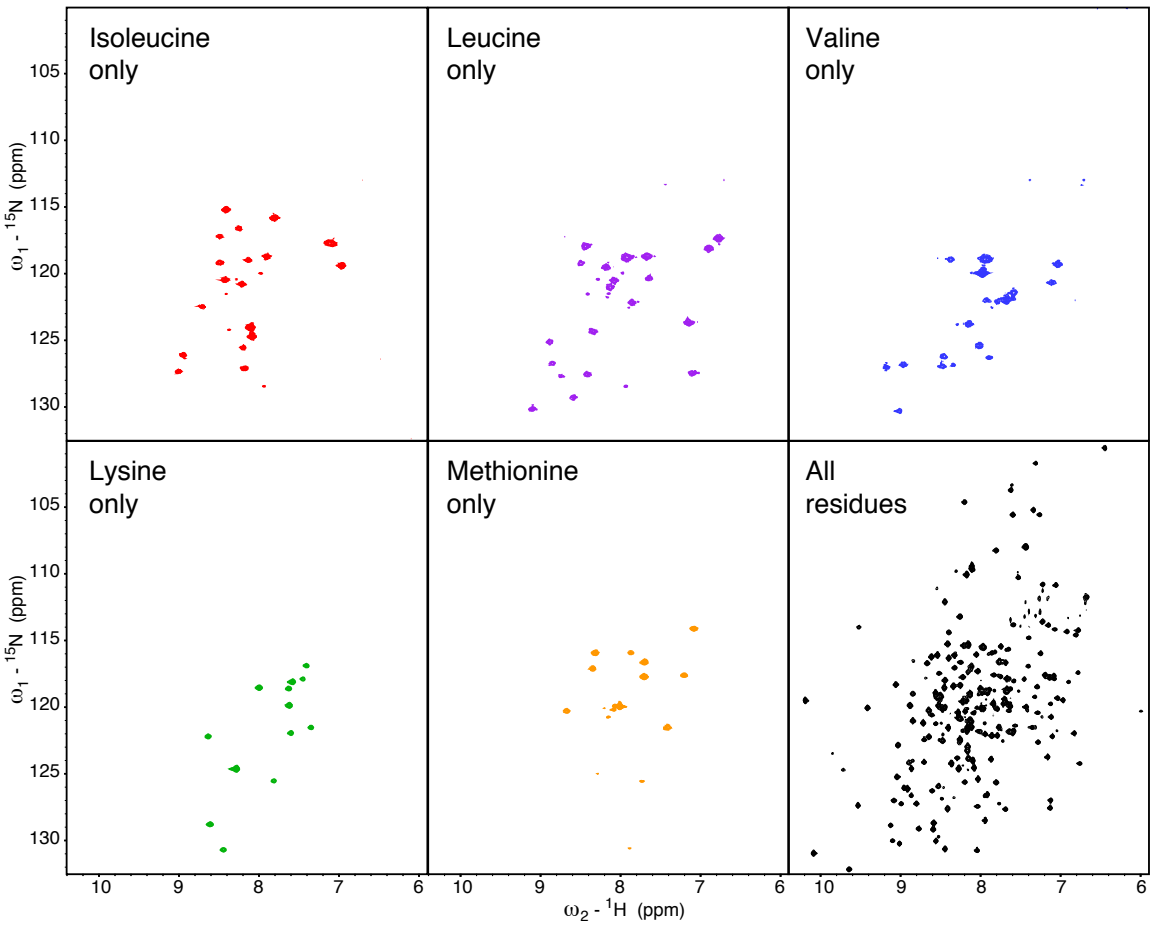

Figure 4.6: Residue-specific labeling of eIF4A-NTD. ^{15}N HSQC spectra are shown of eIF4A-NTD with ^{15}N selectively incorporated for a single amino acid type. From left to right, isoleucine residues are shown in red, leucine in purple, valine in blue, lysine in green, and methionine in orange, followed by a TROSY spectrum of uniformly ^{15}N-labeled eIF4A-NTD.

including Ile, Leu, and Val, and to some degree Phe and Tyr, the transaminases responsible for exchanging the amine group between that amino acid and glutamate, which serves as the main form of ammonia storage, can be efficiently repressed by supplying an abundance of all other unlabeled amino acids. Several other amino acids, including Cys, Met, Lys, His, and Arg, can also be selectively labeled with ^{15}N in ordinary BL21(DE3) e. coli because they occur at the ends of biochemical pathways for amino acid production [66].

Based on the above considerations, and the number of remaining unassigned residues of each type, five amino acids were chosen for selective labeling: Ile, Leu, Val, Lys, and Met. Figure 4.6 shows a ^{15}N HSQC spectrum of each selectively labeled sample, followed by a TROSY spectrum of uniformly ^{15}N-labeled eIF4A-NTD. The isoleucine, leucine and valine samples all show some evidence of scrambling to glutamate, and the valine sample also had noticeable scrambling to alanine, as can be predicted from known biochemical pathways. Some of this scrambling might have been prevented by using shorter expression times.

4.2.4 Completeness of assignment

The amide backbone assignment of eIF4A-NTD is shown in Figure 4.7. All observable backbone resonances have been assigned. This includes the amide N and H for 222 of the 230 non-proline residues in the sequence. As always, the N-terminus is not observable; several other signals are also missing, likely due to line-broadening caused by conformational exchange. For example, two residues in the Walker A motif, or phosphate-binding loop, are likely sampling multiple conformations. Similarly, no signals were observed for N139, but line-broadening weakens the signals for neighboring residues in the sequence. Crystal structures of the free form of eIF4A-NTD either lack electron density or have high B-factors in this region [6, 30], and this site has been shown to be particularly susceptible to proteolysis [38], so this signal is likely missing due to conformational exchange [6].

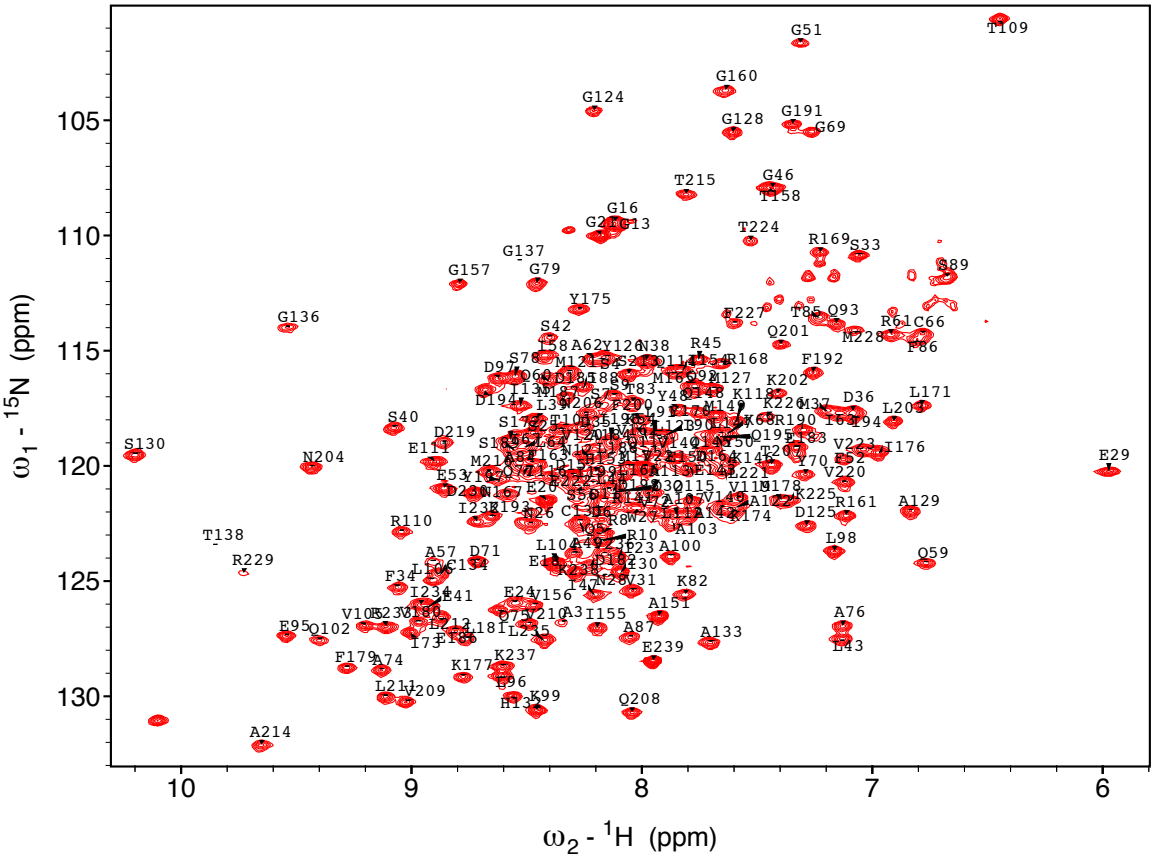

Figure 4.7: ^{15}N TROSY spectrum of eIF4A-NTD with the amide backbone assignment.

4.3 ATP, ADP, and RNA binding surfaces on eIF4A-NTD

Armed with the backbone chemical shift assignment, we titrated ^{15}N-labeled eIF4A-NTD with ATP, ADP, and RNA, probing for the expected interactions with conserved motifs that have been shown to be involved in ATPase and helicase activity in DEAD-box proteins.

ATP induces dramatic chemical shift changes across a large portion of the protein. When adequate magnesium is supplied for tight binding, the interaction appears to be exclusively in the slow exchange regime: peaks in the HSQC representing the free state gradually disappear as the ATP concentration increases, while new peaks representing the bound state gradually appear. We observe the same behavior for ADP binding, although the affinity appears to be lower: a fourfold excess of ADP does not appear to saturate 0.2mM eIF4A-ntd although a fourfold excess of ATP does appear to be approximately saturated. Figure 4.8 shows a ^{15}N TROSY spectrum of eIF4A-NTD in black, overlaid with spectra of eIF4A-ntd in slow-exchange complexes with ADP (magenta) and ATP (cyan). There are several cases where peaks in the ADP-bound spectrum appear to be split between the free state and a bound state which closely matches the ATP-bound peak, most of which appear to map to residues physically near known motifs for ATP binding and helicase activity. These results suggest that eIF4A-NTD binds less tightly to ADP than to ATP, which appears to contradict previous reports [39]. On the other hand, without collecting additional spectra with even higher concentrations of ADP, it is difficult to rule out the possibility that we are indeed observing the fully saturated state, but that the saturated ADP-bound state includes slow conformational exchange that is not present in the saturated ATP-bound state. Alternatively, it is possible that the full-length eIF4A does bind more tightly to ATP, while the N-terminal domain alone binds less tightly, especially since conserved motif VI in the C-terminal domain is likely to interact with the phosphates.

Supplying less magnesium weakens but does not eliminate nucleotide binding, pushing

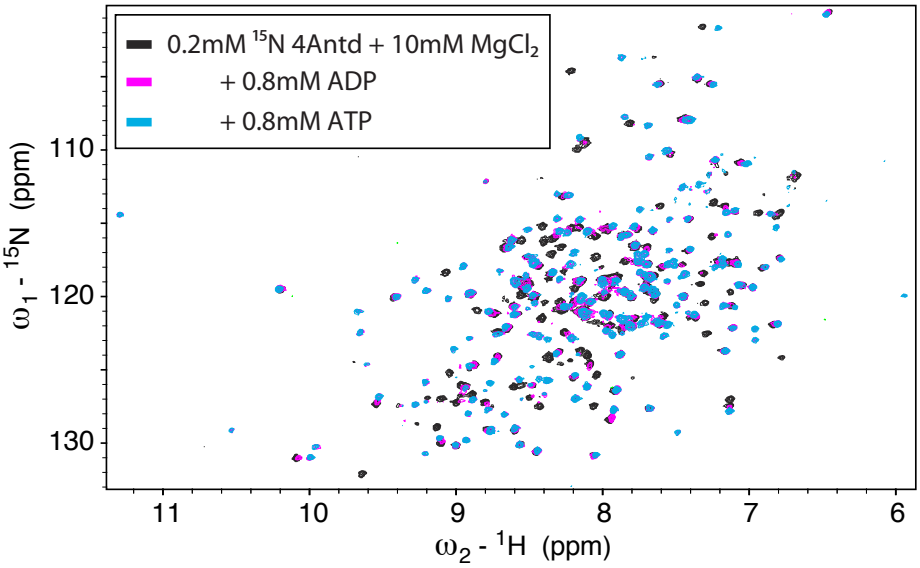

Figure 4.8: Comparison of ^{15}N TROSY of eIF4A-NTD with ATP and ADP. Black contours show free eIF4A-NTD in high magnesium buffer, overlaid with spectra of eIF4A-NTD with fourfold excess ADP (magenta) and fourfold excess ATP (cyan).

some peaks into the fast exchange regime, allowing us to track their movement from the free state to the bound state. Figure 4.9 shows the ^{15}N TROSY spectra of 300μM eIF4A-NTD as it is titrated with increasing amounts of ATP, ranging from 75-600μM. The inset shows the variety of responses peaks in the spectrum have to increasing amounts of ATP. V209 is unaffected by ATP binding, whereas A74 is in intermediate exchange, disappearing altogether before reappearing elsewhere. F179 is in fast exchange, gradually moving upwards and leftwards as the ATP concentration increases, with little change in linewidth. L211 shows an example of moderately fast exchange, with some characteristics of intermediate exchange: the peak shifts as in fast exchange, but also becomes weaker and broader, and then stronger and sharper again as the binding is saturated.

Panel A of Figure 4.11 shows the chemical shift changes that occur when 0.3mM ^{15}N-labeled eIF4A-NTD is titrated with 0.6 mM ATP. The normalized distance in ppm is shown for those peaks that are in the fast-exchange regime and whose movements could be

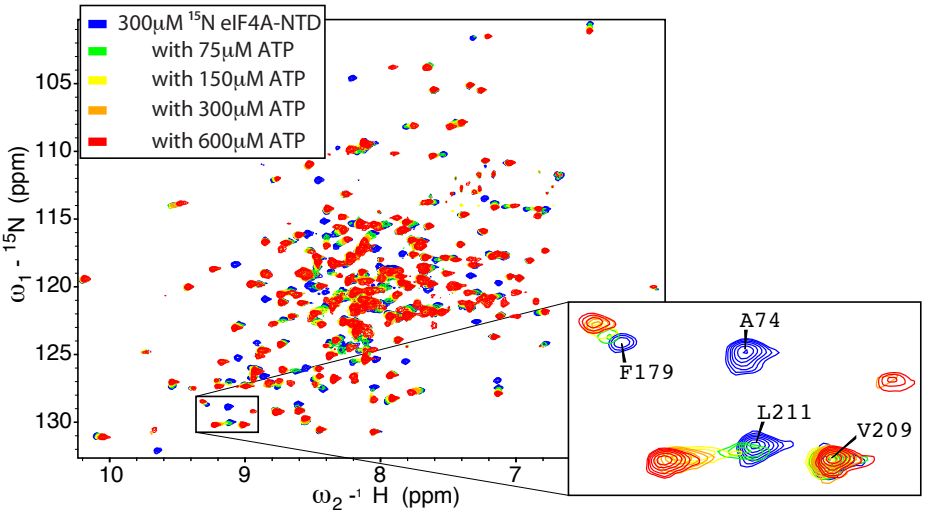

Figure 4.9: ^{15}N TROSY titration of ^{15}N eIF4A-NTD with ATP. 300μM ^{15}N eIF4A-NTD alone is shown in blue contours, while 4:1, 2:1, 1:1, and 1:2 complexes are shown in green, yellow, orange, and red contours, respectively.

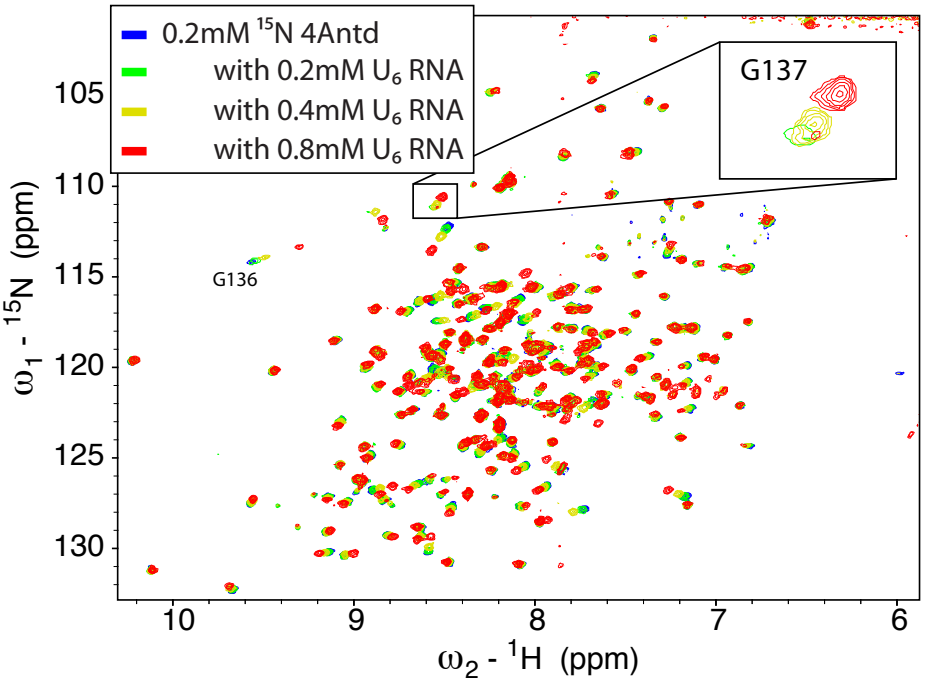

Figure 4.10: ^{15}N TROSY titration of ^{15}N eIF4A-NTD with U_6 RNA. 200μM ^{15}N eIF4A-NTD alone is shown in blue contours, while 1:1, 1:2, and 1:4 complexes are shown in green, yellow, and red contours, respectively.

tracked as the ATP concentration was gradually increased in four steps. For those peaks that are in slow exchange and whose movements could not be tracked, the distance is shown arbitrarily as 0.4 ppm. Many of the strongest changes occur in conserved DEAD-box helicase motifs, which are marked above the charts. This effect is particularly pronounced for the Q-motif (F and GFEKPSAIQ), and the Walker A and Walker B motifs (AQSGTGKT and DEAD, respectively. Significant changes also occur in motifs III and Ia (SAT and PTRELA), though motif Ia is in fast exchange. Interestingly, residues near the interdomain linker of eIF4A also undergo significant changes, raising the possibility that ATP brings the two domains together through some allosteric influence on the linker itself, and not only through interactions with motifs on both sides of the interdomain cleft. This possibility is appealing because the methyl groups of V236 immediately prior to the interdomain linker pack against I58 of the Q motif and A76 of the Walker A motif. Of course, it could simply be the result of a nonspecific charged interaction between the negatively charged phosphates and the positively charged residues at the C-terminal end of the domain. Panel C of Figure 4.11 shows the eIF4A-NTD surface colored in dark blue, oriented to show the eIF4A-CTD interface, with the residues most affected by ATP binding colored in light green.

The N-terminal domain of eIF4A contains the active site for ATP hydrolysis, as well most of the residues implicated in ATP binding, so there has been some concern that ATP hydrolysis could be taking place during the titration. However, it is worth noting that eIF4A is an RNA-dependent ATPase, meaning that very little hydrolysis activity is observed even for the full-length helicase in the absence of RNA, and no RNA was present in these studies. The ATPase activity of full-length eIF4A has also been shown to be strongly dependent on buffer conditions, with relatively low activity in conditions resembling our NMR buffer [39]. While we have not observed significant ATPase activity in the N-terminal domain alone,

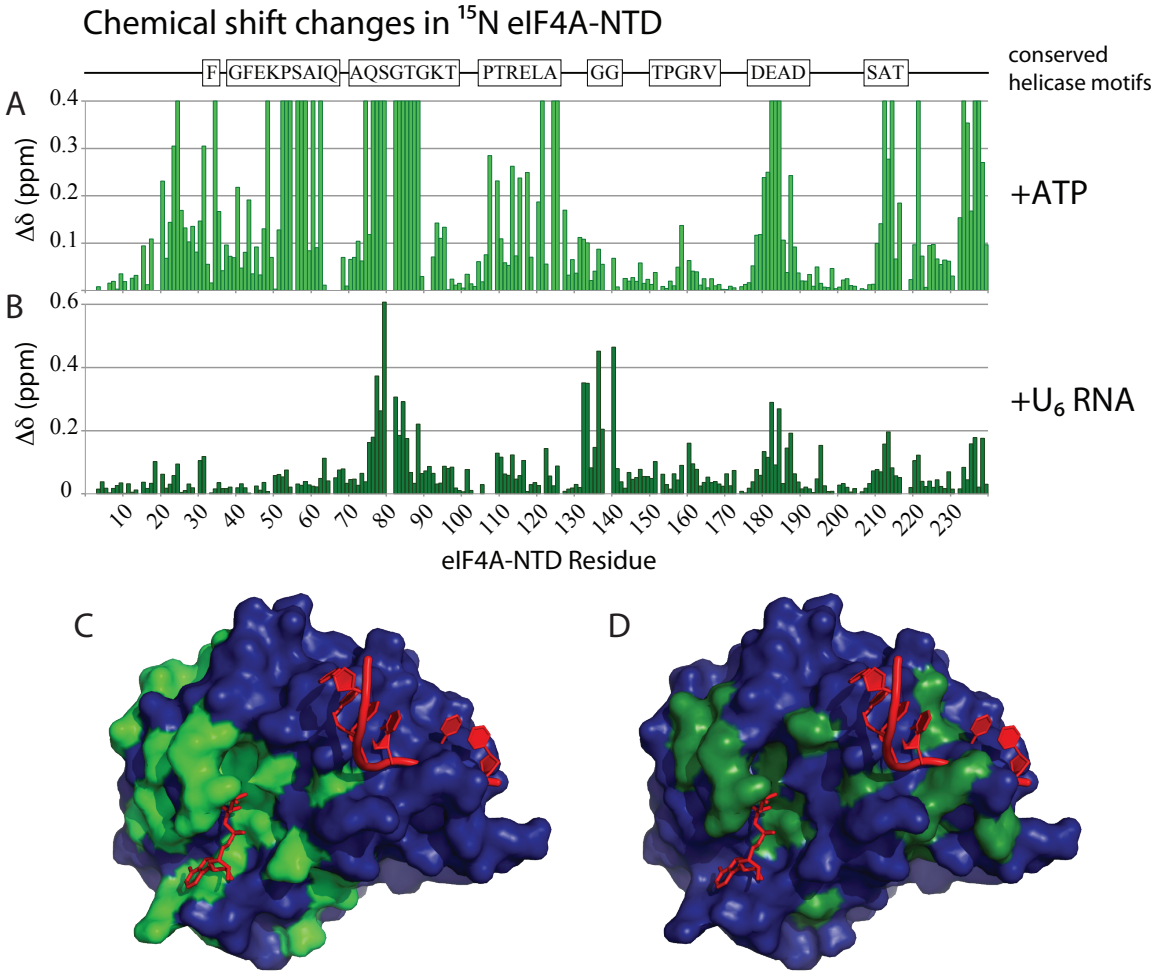

Figure 4.11: Chemical shift perturbations of eIF4A-NTD caused by ATP (A) and RNA (B). Residues whose chemical shift changes could not be tracked due to slow exchange in the ATP titration are shown with a 0.4ppm shift. The conserved DEAD-box helicase motifs are shown in their approximate location above the charts. Panels C and D show eIF4A-NTD (2G9N) colored in dark blue, with residues experiencing the largest shifts in the presence of ATP and RNA colored in light green and dark green, respectively. ATP and RNA are shown as red sticks, positioned as in the structure of the closed conformation of eIF4AIII (2HYI) to indicate the domain orientation, showing the CTD interface. From this perspective, ATP appears in the lower left, and RNA in the upper right of the face of eIF4A-NTD.

others have measured rates as high as 0.2-0.5 pmol/hr for 1 pmol protein for eIF4A-NTD in conditions meant to resemble the cytosol [6].

Because eIF4A has little known sequence specificity for RNA binding, we opted to use a short, simple RNA sequence, also hoping to minimize the signal loss that occurs when higher molecular weight complexes are formed. The crystal structure of the homologous eIF4AIII protein in complex with RNA suggests that a 6-mer is sufficient to cover the binding site. We chose to use the same RNA sequence, U_6, but also performed the same titration with other sequences, and observed the same shifts.

RNA binding affects the ^{15}N TROSY spectrum of eIF4A-NTD in surprising ways. As expected, significant changes occur in the region around the GG motif. The spectra are shown in 4.10, with G136 and G137 of the GG motif labeled. G137 is highlighted in the inset because the signal is very weak in the free form, likely due to line-broadening caused by conformational exchange, yet the signal grows stronger upon RNA binding, as if the RNA stabilizes a single conformation. This interpretation seems inconsistent with reports that RNA binding increases proteolytic susceptibility at this site[38]. Based in their report, it would seem that the G137 peak coalesces into a single, sharp peak by becoming more mobile, changing from intermediate time-scale conformational exchange to fast conformational exchange. More likely, the GG motif in particular becomes less mobile, while the subsequent residues remain relatively mobile, yet somehow become more suitable targets of proteolysis.

Strangely, RNA also appears to interact with the Walker A and Walker B motifs normally responsible for coordinating the phosphates of ATP, as shown in the bar chart in panel B of Figure 4.11. In principle, the chemical shift changes in the ATP binding site could be caused by allosteric changes associated with RNA in the normal binding site, but the effect is large enough to suspect that the phosphate-coordinating motifs are interacting with the

phosphate backbone of the RNA directly. Again, the most affected residues are painted in dark green on the otherwise dark blue surface of eIF4A-NTD in panel D of Figure 4.11. In this panel and panel C, ATP and RNA are shown as red sticks, positioned as in the structure of the closed conformation of eIF4AIII (2HYI) [2]to indicate the domain orientation.

4.4 HEAT2 BINDING INTERFACE ON EIF4A-NTD

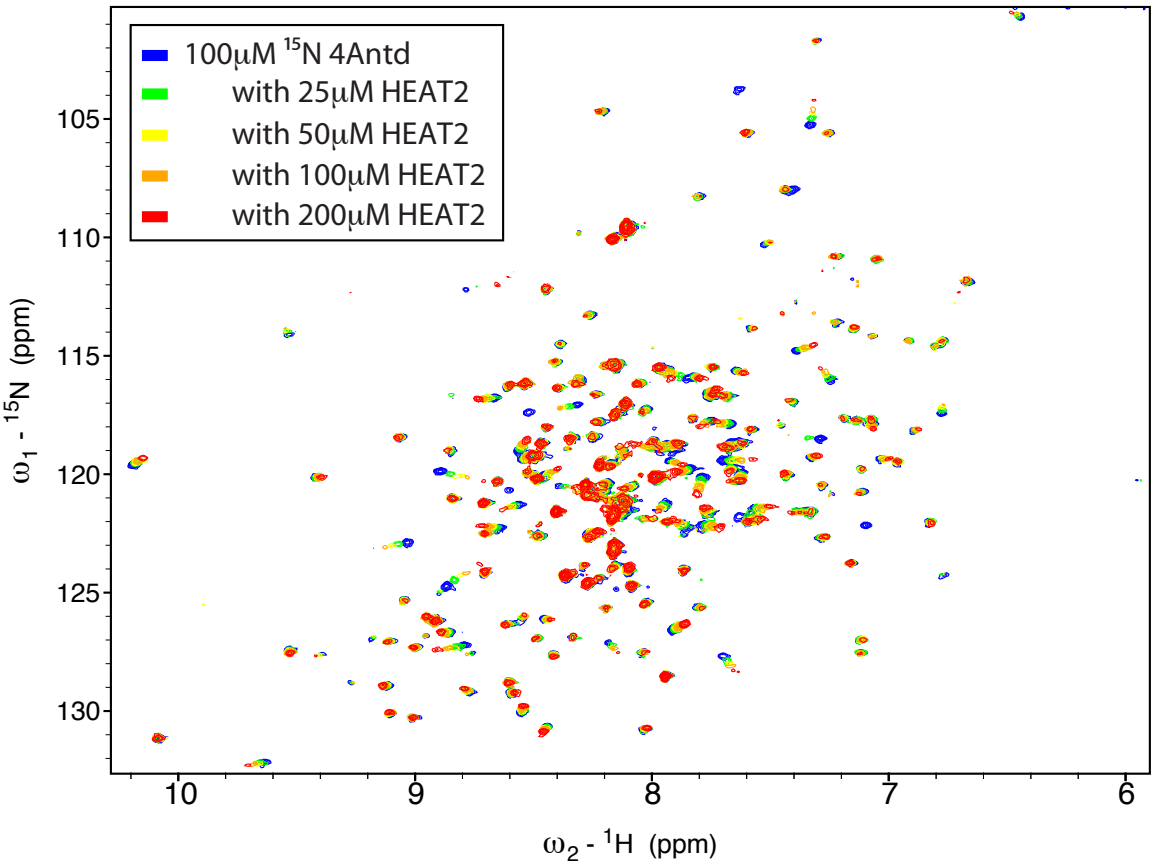

Figure 4.12: ^{15}N TROSY titration of ^{15}N eIF4A-NTD with unlabeled HEAT2. 100μM ^{15}N eIF4A-NTD alone is shown in blue contours, while 4:1, 2:1, 1:1, and 1:2 complexes are shown in green, yellow, orange, and red contours, respectively.

Finally, we look at ^{15}N-labeled eIF4A-NTD in complexes with other protein domains,

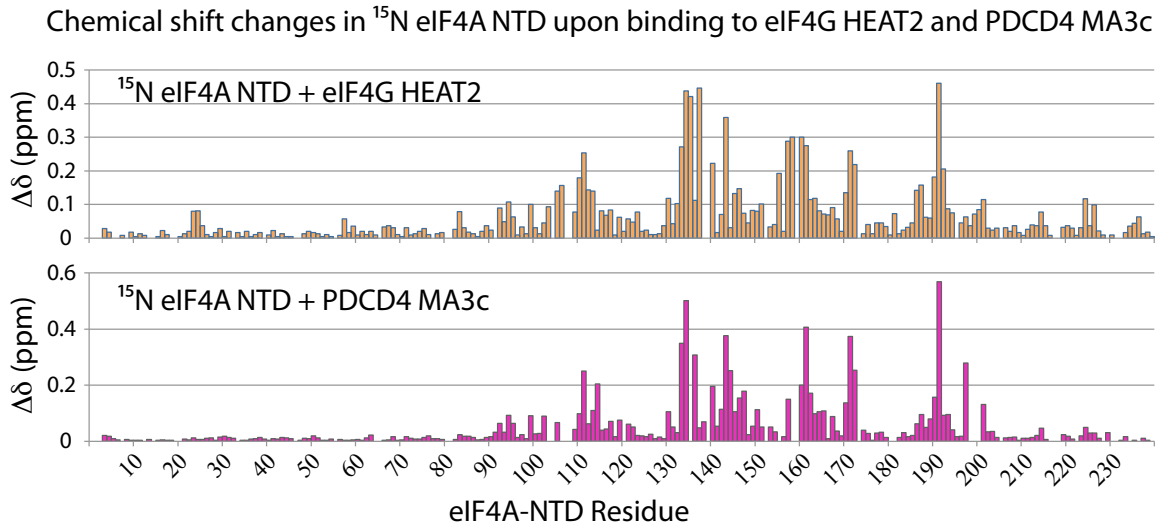

Figure 4.13: eIF4G HEAT2 and Pdcd4 MA3c share a common interface on eIF4A-NTD. The top panel shows the chemical shift changes caused by titration of 200μM unlabeled eIF4G HEAT2 into 100μM ^{15}N eIF4A-NTD, while the bottom panel shows the chemical shift changes caused by 400μM unlabeled Pdcd4 MA3c in 200μM ^{15}N eIF4A-NTD.

including eIF4G HEAT2 and the MA3c domain of Pdcd4. Figure 4.12 shows the ^{15}N TROSY spectra of 100μM eIF4A-NTD as it is titrated with increasing amounts of unlabeled HEAT2, ranging from 25-200μM. Most peaks that shift significantly upon binding are in the fast exchange regime, and are easily tracked. In most cases, peak intensities decrease, partly due to formation of a larger, slower tumbling complex, and the resulting faster relaxation. In many cases, some degree of intermediate exchange contributes to additional decrease in peak intensity. Figure 4.13 shows the chemical shift changes that occur in eIF4A-NTD when titrated with unlabeled HEAT2 domain in the top panel. Several peaks appear to shift in the slow or intermediate exchange regime or to vanish entirely upon binding, including residues 157-161. Where possible, guesses were made about where the peaks reappear in the spectrum. The final positions of peaks for residues 158 and 160 could not be found. Since they are strongly affected, an arbitrary value of 0.3 ppm was assigned for data display purposes, approximately matching the neighboring residues in the chart.

The bottom panel of Figure 4.13 shows the chemical shift changes that occur in eIF4A-NTD in the presence of the unlabeled MA3c domain of Pdcd4. The pattern is strikingly similar to the one above for HEAT2, as expected based on sequence homology and structural similarity, as well as mutagenesis data that show that critical residues for binding to eIF4A are conserved across HEAT2 and both Pdcd4 MA3 domains [70]. We have also confirmed by NMR that these domains compete for binding to eIF4A [60].

Figure 4.14 shows eIF4A-NTD painted in dark blue, with the residues most affected by HEAT2 binding painted orange, from two perspectives. In the left panel, eIF4A-NTD, is shown as a ribbon, positioned with the eIF4A-CTD interface at the top. In the middle panel, ATP and RNA are shown, positioned as in the closed structure of eIF4AIII [2]. The middle panel clearly shows that the HEAT2 binding site and the RNA binding site overlap significantly, suggesting that the HEAT2 domain competes with RNA for binding to eIF4A.

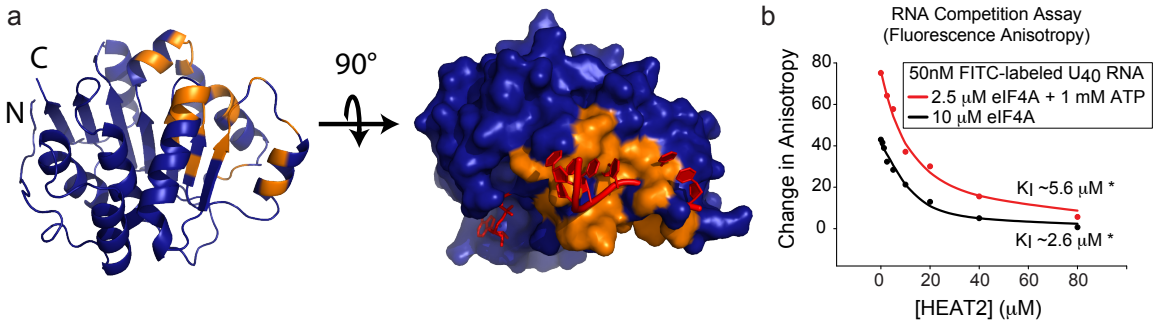

Figure 4.14: eIF4A-NTD (PDB 2g9n) is shown in dark blue with the residues most affected by HEAT2 binding painted orange (a). The center panel shows a surface representation of the HEAT2 binding site on eIF4A-NTD, with RNA and ATP modeled as they are positioned in the crystal structure of the homologous eIF4AIII. Panel b contains fluorescence anisotropy graphs showing the ability of eIF4G HEAT2 to inhibit binding of 10μM eIF4A to 50 nM FITC-labeled U_{40} RNA in the absence of nucleotide (black) and 2.5μM eIF4A to 50 nM FITC-labeled U_{40} RNA in the presence of 1mM ATP (red).

We sought to confirm that HEAT2 can displace RNA from eIF4A-NTD directly by NMR of ^{15}N eIF4A-NTD, but our efforts were complicated by the ability of RNA to bind to the ATP binding site: spectra show that RNA and HEAT2 appear to be able to bind

simultaneously to eIF4A-NTD, but probably not in a way that is physiologically relevant. Instead, we used fluorescence anisotropy to confirm competition between RNA and HEAT2 on full-length eIF4A, in the presence and absence of ATP. We measured the change in fluorescence anisotropy of FITC-labeled U_{40} RNA upon addition to a complex of eIF4A and HEAT2, at various HEAT2 concentrations. Figure 4.14b shows that HEAT2 competes with RNA binding to eIF4A with a K_I of approximately 2.6μM in the absence of ATP, and is a less effective inhibitor of RNA binding in the presence of ATP.

4.5 Additional interactions and negative results

Our original model of the eIF4A-eIF4G interaction (Figure 1.6) led us to hypothesize that the third HEAT domain of eIF4G (HEAT3) might also contact eIF4A. Such an interaction would be weak, since it was not detected by the previous experiments with deletion mutants that discovered and located the binding sites in the HEAT1 and HEAT2 domains, but chemical shift perturbation assays are much more sensitive to weak binding than the pull-down type of experiments that were previously used. However, we were unable to observe any changes in spectra of ^{15}N-eIF4A-NTD upon addition of HEAT3 alone, and a construct containing both the HEAT2 and HEAT3 domains caused no additional changes, compared to the HEAT2 domain alone. Therefore, we conclude that the HEAT3 domain of eIF4G does not also participate in binding to eIF4A.

We have been unable to observe any chemical shift perturbations in ^{15}N eIF4A-NTD when titrated with either eIFG HEAT1 or eIF4H, despite our expectation of a biologically relevant interactions. In both cases, for HEAT1 and eIF4H, the interaction with full-length eIF4A is considerably strengthened by the presence of ATP, which was omitted from our chemical shift perturbation assays. We believe that the interactions require not only that loops within the N-terminal domain must be in the ATP-bound conformation, but also

that the two domains of eIF4A must be brought together by the ATP, to form a continuous binding surface for HEAT1 or eIF4H.

4.6 Orienting HEAT2 on eIF4A-NTD

4.6.1 Paramagnetic Relaxation Enhancements

Site-directed spin labeling and measurement of paramagnetic relaxation enhancement (PRE) effects can be useful for obtaining distance constraints when it is difficult to obtain sufficient NOE data, which can occur when an interface is predominantly α-helical and polar, as in the case of HEAT2 and eIF4A-NTD, and especially when a complex is large enough that extensive deuteration is required. PREs are especially suitable for intermolecular restraints, since they are sensitive over much longer distances than NOEs, typically most sensitive over 15-25Å, rather than 5Å. The technique typically involves the incorporation of a paramagnetic spin label such as MTSL, which has a nitroxide free radical (Figure 4.16, left panel) into a protein via reaction with a single solvent-exposed cysteine. The paramagnetic nitroxide causes distance-dependent line-broadening of resonances of nearby nuclei. The diminished peaks can be restored by adding ascorbate to reduce the nitroxide to the corresponding hydroxylamine (Figure 4.16, right panel). A comparison of peak intensities before and after reduction by ascorbate (Figures 4.17 and 4.18) indicates which residues are nearest to the spin label.

The relative orientations of the HEAT2 domain and eIF4A-NTD in the binary complex were determined using PREs with site-directed spin labeling, following an approach similar to that employed by Battiste and Wagner [4]. Site-directed mutagenesis was used to prepare a panel of six single-cysteine mutants, with mutations distributed around the proposed HEAT2 binding site. Figure 4.15 shows the sites for attaching the spin label, depicted as yellow spheres, surrounding the HEAT2 binding site on eIF4A-NTD, which is painted

orange on the dark blue background. The nitroxide spin label MTSL was attached at each of these sites by formation of a disulfide bond with the single cysteine. Some effort was made to avoid mutating conserved motifs (Figure 1.3), in the hopes of preserving a functionally active and relevant structure. For this reason, no cysteine mutation was made to residues 78-80 in the Walker A motif, located directly above the orange patch in the right panel of Figure 4.15, although the position would likely have been useful for providing distance restraints.

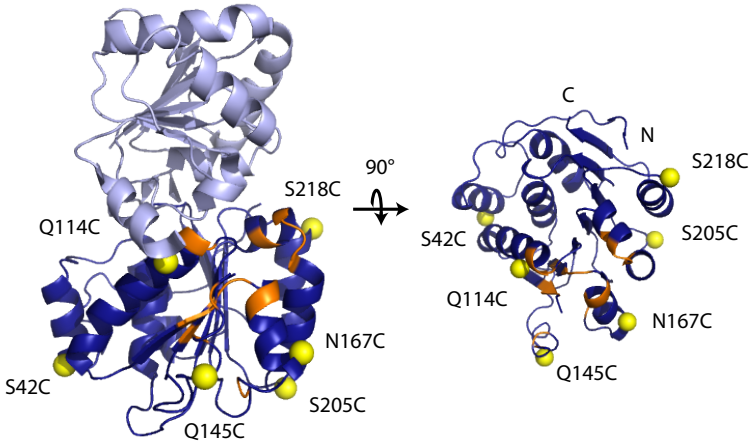

Figure 4.15: Sites of single-cysteine mutations for attaching a spin label are shown as yellow spheres on eIF4A-NTD (dark blue). Residues in eIF4A-NTD implicated in binding to HEAT2 are painted orange. The left panel provides perspective with the CTD (light blue) positioned as in eIF4AIII (2HYI), and the right panel shows the view of the purported CTD interface.

^{15}N TROSY spectra were recorded on ^{2}H^{15}N HEAT2 in complex with each spin-labeled mutant of eIF4A-NTD, before and after reducing the nitroxide with ascorbic acid. A representative pair of spectra of an oxidized and reduced complex is shown in Figure 4.17. Several peaks have vanished entirely, and many others are significantly diminished. The inset of Figure 4.17 highlights some of these peaks. Many of the most strongly affected peaks belong to residues already known to be in the binding site, based on chemical shift changes. Others belong to nearby residues on neighboring HEAT repeats.

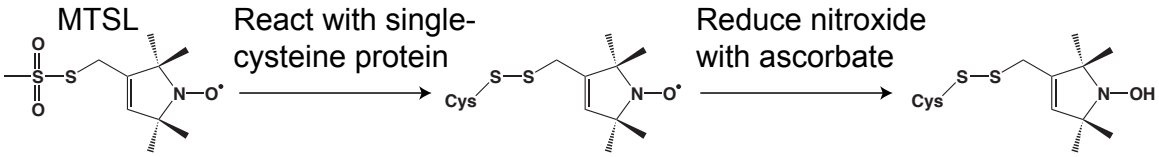

Figure 4.16: MTSL is attached to a single-cysteine mutant by formation of a disulfide bond, generating a spin-labeled protein in the oxidized form. After the first spectrum is recorded, the nitroxide spin label is reduced by the addition of ascorbic acid. Another spectrum is then recorded of the reduced form.

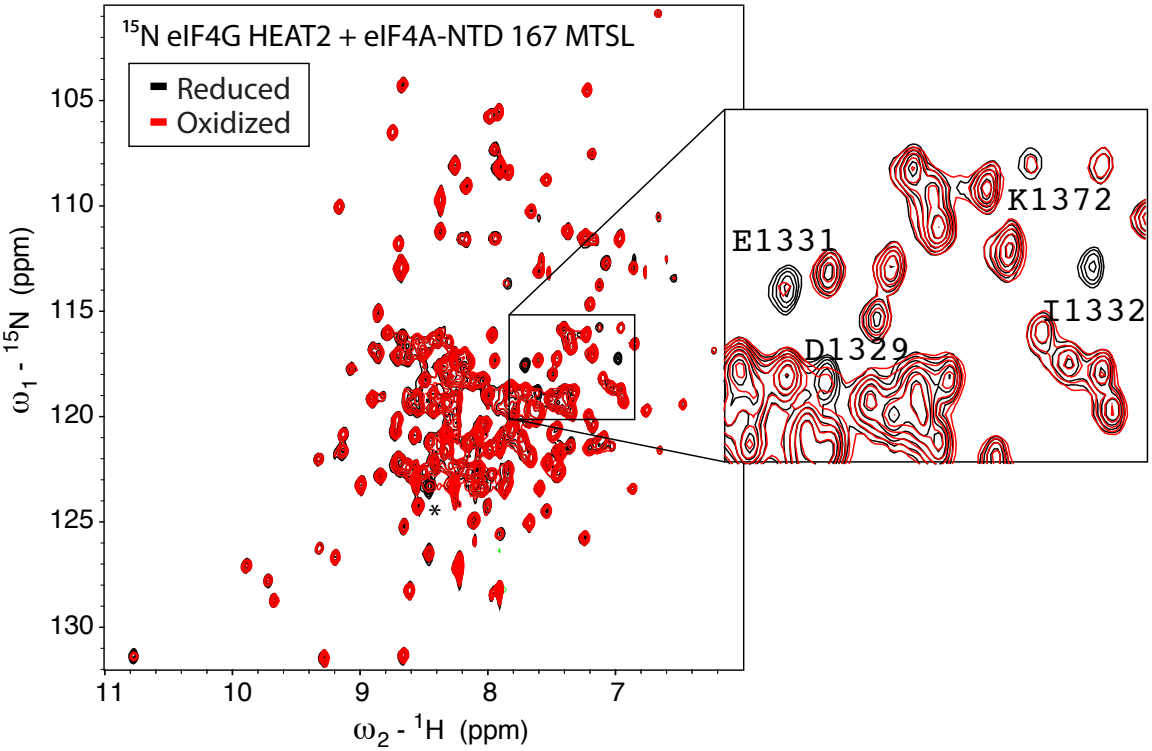

Figure 4.17: Representative ^{15}N TROSY spectra of eIF4G HEAT2 in complex with eIF4A-NTD N167C MTSL. The spectrum of the reduced form is shown in blue, and the oxidized form is shown in magenta. The inset highlights several residues with significant relaxation enhancement by the paramagnetic spin label.

As expected, no significant changes were observed between the spectra taken before and after reduction of the nitroxide spin label for samples with the MTSL attached at sites 42 and 205. A bar chart of the intensity ratios $\left(\frac{oxidized}{reduced}\right)$ for each peak in the HEAT2 spectrum with eIF4A-NTD-205C is shown in the top panel of Figure 4.18 to illustrate several pitfalls of these experiments.

If residue 205 is indeed too far from the binding site for the nitroxide spin label to influence amide proton relaxation rates in the HEAT2 domain, then the oxidized and reduced spectra should be identical. For every residue in HEAT2, the ratio of the peak heights in the two spectra should be exactly 1. Instead, most peaks have nearly twice the intensity in the oxidized spectrum as in the reduced spectrum. One possible cause for such a consistent change in peak intensities can be precipitation of the ^{15}N-labeled protein. Indeed, precipitate did accumulate in the NMR tube over the course of the experiment, but eIF4A-NTD is far more prone to precipitation than HEAT2, and the single-cysteine mutants even more so. Moreover, residues 1231-1235 and 1428-1437 do have peak height ratios that are nearly 1; they appear to be present and soluble in the same concentration in both experiments. These are the extremely flexible residues at the N- and C-termini of the domain, with very long T2s. More likely, HEAT2 is not precipitating, but forming some kind of soluble oligomers, with slower tumbling and shorter T2 times causing the reduced peak intensities. Oligomerization of the rigid domain would be expected to have little effect on the tumbling rate of the flexible tails. Perhaps the exposed cysteine sidechains on HEAT2 form intermolecular disulfide bonds over the course of the experiment, since the usual reducing agent was omitted from the sample buffer in order to avoid reducing the nitroxide prematurely. This problem was less pronounced in subsequent experiments when lower protein concentrations were used and more care was taken to degas sample buffers thoroughly. In several cases, it was necessary to compensate for high ratios by normalizing to 1, in order to obtain

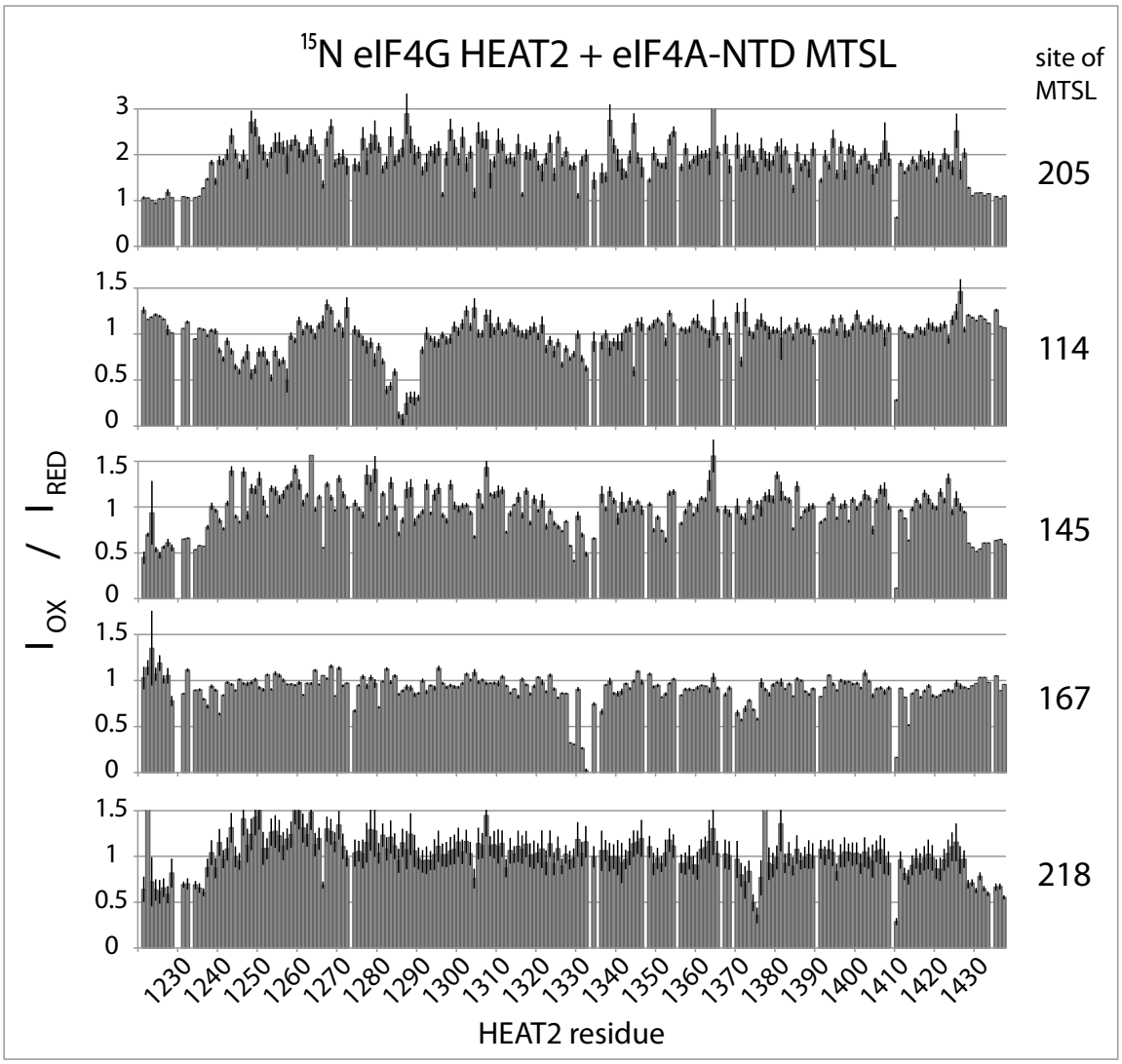

Figure 4.18: Ratios of peak intensities in the spectra of HEAT2 with oxidized and reduced eIF4A-NTD-MTSL, for each of five sites on eIF4A-NTD. Error bars show one standard deviation, as calculated from the noise levels in each spectrum. Bars are missing for all residues which do not appear in $^{1}H^{15}N$-TROSY spectra of HEAT2 with 4A-NTD, including D1333 and all proline residues.

reasonable values for converting to distance restraints.

Another general issue worth noting in the top chart of Figure 4.18 is that residue 1410 has a particularly low peak intensity ratio, as if it is near the nitroxide spin label, despite being surrounded by unaffected peaks. The peak nearly vanishes in the oxidized spectrum, but is actually quite strong in the reduced spectrum, as can be seen in Figure 4.17, marked by an asterisk. This same residue appears to be affected by MTSL in every experiment, irrespective of where the spin label is attached on eIF4A-NTD. This kind of effect has previously been observed by Battiste et al., who proposed that there might be a hydrophobic patch on the protein surface, for which MTSL itself has some affinity, and the resulting weak, transient interaction can affect a single peak [4].

The remaining charts in Figure 4.18 show the peak height ratios for each residue in HEAT2 when paired with eIF4A-NTD with MTSL attached at residue 114, 145, 167, and 218. Dimerization appears to have been a problem when samples for residue 145 and 218 were measured, but but not for samples 114 and 167. A qualitative inspection of the data reveals and approximate interdomain orientation: The spin label attached at position 114 affects residues near the N-terminus of HEAT2, but at position 167, it affects mostly residues in the turn of the third HEAT repeat, and also a few residues closer to the C-terminus.

4.6.2 Docking

We searched for possible docking models of the complex of HEAT2 and eIF4A-NTD using TREEDOCK software [13], while incorporating the experimental results from chemical shift mapping and PRE experiments. The initial structural coordinates used for docking were taken from crystal structures: PDB ID 1UG3 for HEAT2, and PDB ID 3EIQ for eIF4A-NTD. We chose to use 3EIQ, which contains the coordinates of eIF4A bound to PDCD4, instead of free eIF4A-NTD from 2G9N, with the hope that the flexible loops in

the interface with HEAT2 might adopt a conformation similar to their positions in complex with the homologous PDCD4 domains than to the free form. Ultimately, we can test this assumption by performing the same procedure with the other starting structure, but that may not be necessary because we hope to incorporate additional conformations of flexible loops in the interface, by exhaustive search.

The initial phase of docking generates all possible interdomain orientations without steric clashes. The search space is limited by only generating complexes that have at least one point of contact between residues that have been experimentally determined to be part of the binding site on each side of the interface. These anchoring residues were selected from the chemical shift mapping data shown in Figures 4.14 and 4.3.

This initial docking step generates a very large and diverse set of possible interdomain orientations. Models were selected from this set which satisfy experimental distance restraints from PRE data. At this stage, a simplified set of restraints was used, with a single upper limit restraint for all residues with peak height ratios significantly under 1. The chosen cutoff was 33Å from the amide nitrogen of HEAT2 to the Cβ of eIF4A-NTD, based on 25Å between the nitroxyl oxygen of MTSL and the affected amide nitrogen, and 8Å between Cβ and the nitroxyl of MTSL. This distance-based filtering resulted in a set of 15 closely related structures, with a more precisely defined contact surface. A representative model is shown in Figure 4.19, with sites for spin label attachment shown as spheres.

The next step of the docking process will involve energy minimization of the geometrically reasonable structures, by repeating the exhaustive search process with the more narrowly defined interface, as well as additional conformations of short flexible loops and sidechains in the interface, which will be remodeled using OCTOPUS [14]. The model we have now could be improved with additional restraints from new spin label sites, which we can select with the better understanding we now have of the relative orientations of the two domains. With

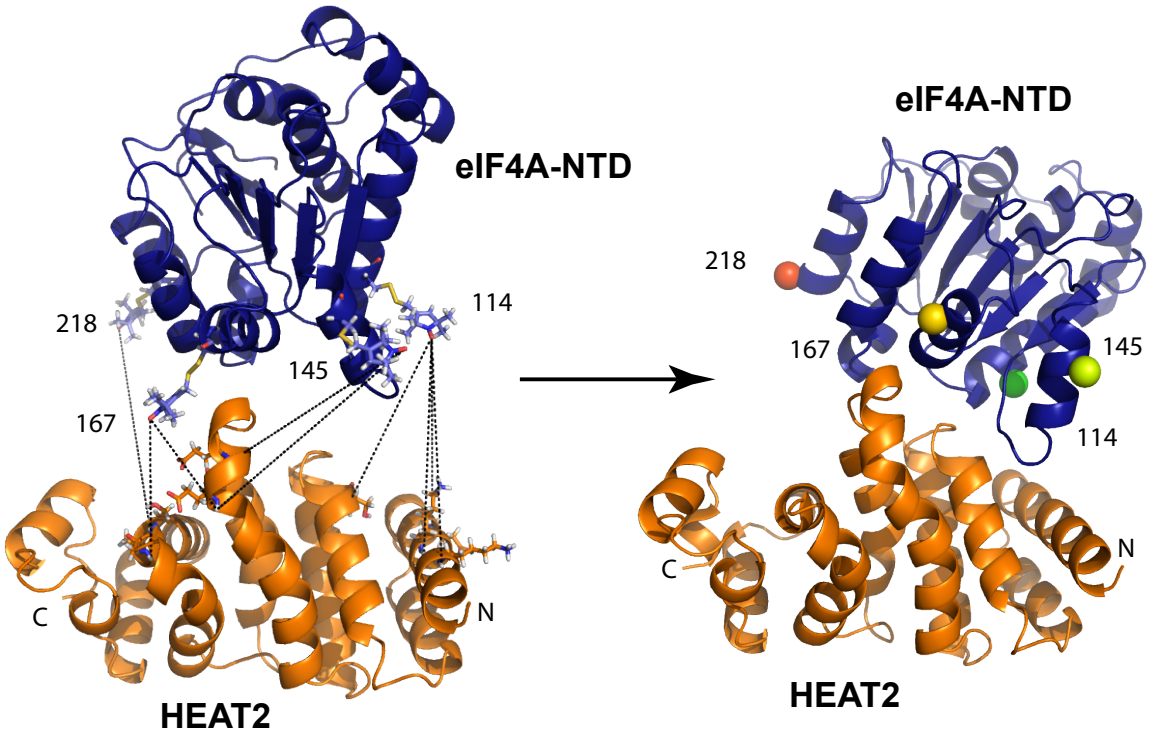

Figure 4.19: Docking eIF4A-NTD with HEAT2. The left panel shows the two domains as they are oriented in our starting model (Figure 1.6), with the spin label MTSL shown as sticks at each of the four mutation sites from which distance restraints were obtained. Several of these restraints are shown as dashed lines. The right panel shows a representative model after docking according to chemical shift perturbations and culling by PRE distance restraints. Sites where the spin label MTSL was attached are shown as colored spheres.

the current restraints, the position of eIF4A along the long axis of the HEAT2 domain is relatively well-restrained by the high-quality data obtained from MTSL attached at residues 114 and 167. However, in the opposite direction, around the long axis, restraints are inadequate. This is partly because we avoided mutating residues in the interface between eIF4A-CTD and eIF4A-NTD, and also because residues like 145 and 218 have turned out to be quite far from the interface, and provide very few restraints. We plan to attach MTSL to residue 78, with the hopes that it is close enough to HEAT2 in the complex to alleviate some of this uncertainty in domain positioning around the long axis. The model also suggests that sites such as 139, 183, and 194 might provide more usable restraints than 145 and 218.

4.7 Conclusions

Chemical shift perturbations confirm that HEAT2 binds to eIF4A in a mode that is very similar to how each MA3 domain in Pdcd4 binds. These observations are consistent with their common ability to compete with RNA and with each other for binding to eIF4A, as well as with the common sites for mutations that prevent binding. However, all of these observations about their similarity fail to address a critical question of why their ultimate effect on translation appears to be so different. Pdcd4 has an unambiguously inhibitory effect on cap-dependent translation initiation, including not only competing with RNA binding, but also interfering with duplex unwinding [69] and ATPase activity [37].

Meanwhile, eIF4G HEAT2 interferes with ATP binding, ATPase activity [33], and RNA binding, yet in total appears to have a positive effect on translation. Point mutants of eIF4G that prevent HEAT2 binding to eIF4A have been shown to decrease formation of 48S ribosomal complexes [44]. There are many possible explanations for this difference. First, Pdcd4 has other features which may enable it to sequester eIF4A into inactive complexes, and the simple function of interfering with Pdcd4 binding may therefore stimulate

translation initiation. Another explanation may be that the HEAT2 domain helps to recruit the stimulatory HEAT1 domain to eIF4A. Although the two domains compete for binding to eIF4A in when supplied in trans [42] (Figure 5.11), the net effect may be in an increase in affinity when supplied in cis. This explanation is consistent with the finding that eIF4A is not a stable part of the eIF4F complex in yeast, where eIF4G lacks the second and third HEAT domains. The HEAT2 domain in human eIF4G may provide just enough extra affinity to keep eIF4A stably in the eIF4F complex. There may also be a structural explanation for HEAT2 being less inhibitory than Pdcd4. The translational suppressor may be able to force a greater separation between the domains of eIF4A than HEAT2 can, even when only one of the two MA3 domains is supplied. Although the interaction surfaces are very similar, they are not identical, and there may be a slight shift in the surface, which causes a significant change in the interdomain angle.

Our preliminary docking results hint that the orientation of eIF4A-NTD on HEAT2 may be slightly different from the orientation of eIF4A-NTD on Pdcd4 MA3m, in a direction that would cause smaller separation between the domains. For example, residue 218 in eIF4A-NTD is too far from residue 1374 and 1375 in HEAT2 in the model based on Pdcd4 binding to produce the PRE effects that we observe in solution. More data are needed to improve the reliability of our model, especially since the data collected with MTSL attached at residue 218 are suspect due to the dimerization problems, but we should remain alert to the possibility that HEAT2 may act as a less effective wedge between the eIF4A domains than Pdcd4 does. Figure 4.20 shows our docked model alongside the model generated by replacing Pdcd4 MA3m in 2ZU6 with HEAT2 [9], in approximately similar orientations. The colored spheres representing residues 114, 145, 167, and 218 help to illustrate the different binding angle in the two models.

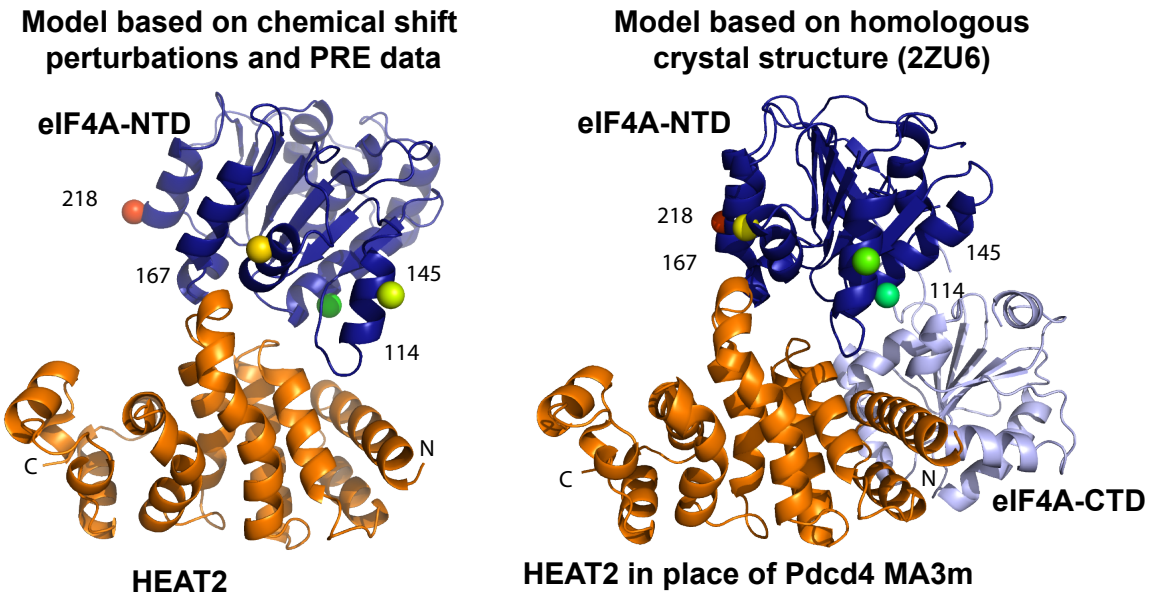

Figure 4.20: Docked model of eIF4A-NTD with HEAT2, based on chemical shift perturbation mapping and PRE data (left panel) has similar, but not identical, domain orientation as in the model of HEAT2 with full-length eIF4A-NTD based on the crystal structure (PDB ID 2ZU6) of eIF4A with Pdcd4 [9] (right panel). The right panel also shows residues 114, 145, 167, and 218 as colored spheres as an orientational reference.

5

Structural characterization of larger complexes involving eIF4A

In this chapter, we combine the results of prior chapters with previously solved structures and known homology to model larger complexes of eIF4A and eIF4G. First, we explore the possibility of expanding upon the two-domain NMR structure of eIF4G-HEAT2 in complex with eIF4A-NTD to work on a complex containing full-length eIF4A. Although the three-domain complex is quite large for NMR work, at 70 kDa, it is surprisingly stable, with tight binding at low salt concentrations. The structures of all individual domains

have previously been solved, and we plan to treat them as rigid bodies, and focus on the interactions among the domains. Next, we describe preliminary work using fluorescence resonance energy transfer (FRET) to measure the distance between the two domains of eIF4A under various conditions, as it changes between closed and open states as proposed in Figure 1.4.

Finally, we refine the original model for the organization of the eIF4A/eIF4G/eIF4H complex, based on the chemical shift perturbation mapping described in Chapter 4, along with other mapped interaction surfaces. This model leads to a new theory for how the scanning preinitiation complex is organized. Next, we address some questions about competitive binding within the complex, and we show evidence that the ternary complex actually exists, despite the competition. We present a model for how the competition within the stable complex can give rise to the dynamic functional behavior of the unwinding machine.

5.1 NMR studies on the complex of full-length eIF4A with HEAT2

NMR studies on full-length eIF4A are challenging due to its tendency to aggregate at relatively low concentrations, making it difficult to study binding events that require both domains. The finding that eIF4A binds tightly to the HEAT2 domain, with a dissociation constant below 1μM under moderate salt conditions (Figure 5.11), raised the possibility of obtaining a stable complex. Size-exclusion chromatography on the complex reveals that the interaction is sensitive to salt concentrations, and can be disrupted with high salt concentrations, but the proteins elute together under low salt conditions (Figure 5.1).

5.1.1 Sample preparation

The stable complex of eIF4A with eIF4G-HEAT2 is far less prone to aggregation than eIF4A alone, permitting us to use much higher sample concentrations than with eIF4A

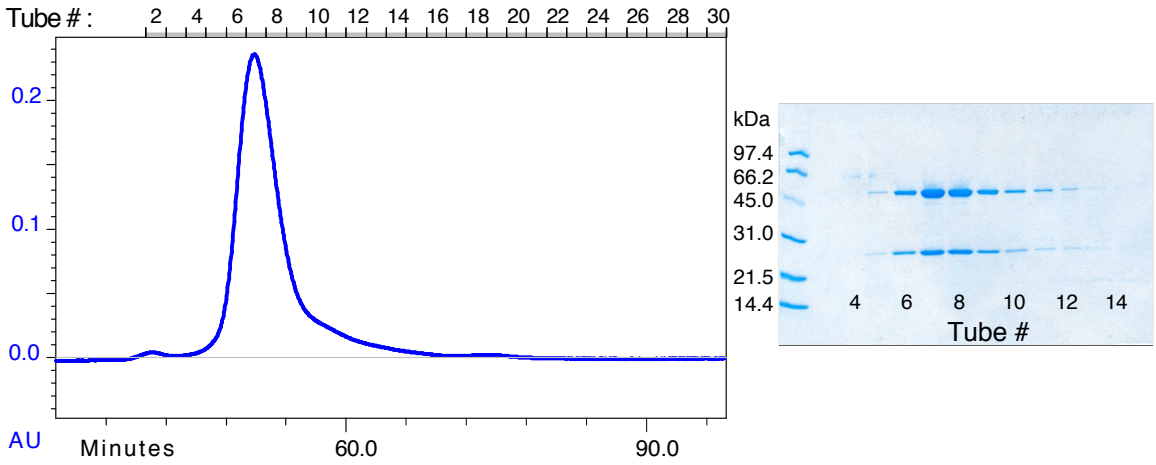

Figure 5.1: eIF4A and eIF4G HEAT2 form a stable complex that elutes as a single peak from a size exclusion chromatography column, as shown by the single peak in the UV trace in the left panel. The SDS-PAGE gel in the right panel shows that each fraction from the peak contains both eIF4A and HEAT2, in a consistent ratio. The upper bands, near 45 kDa are full-length eIF4A, and the lower bands, near 25 kDa, are HEAT2.

alone, and with much less sample precipitation. Moreover, the complex is especially stable at very low salt concentrations, which results in better NMR sensitivity than the high salt concentrations required to prevent precipitation of eIF4A. These qualitative observations of improved sample stability are consistent with circular dichroism (CD) measurements that other researchers have made more recently. Fujita et al. used CD spectroscopy to monitor the unfolding temperatures of eIF4A and its domains, along with the eIF4G HEAT domains, and the complex of eIF4A with a tandem HEAT1-HEAT2 construct [17]. They found that full-length eIF4A unfolds at 44°C, whereas its individual domains are stable up to 46-47°C, and the complex with this segment of eIF4G has a melting transition temperature of 51°C.

Figure 5.2 shows spectral changes in eIF4A and HEAT2 and caused by complex formation. The left panel shows the ^{15}N TROSY spectrum of 100μM eIF4A alone (black contours), overlaid with a ^{15}N TROSY spectrum of the 1:1 complex of ^{2}H^{15}N eIF4A with unlabeled HEAT2, at a 500μM concentration (red contours). The difference in concentration is due to

the susceptibility of free eIF4A to aggregation at the high concentrations used for NMR, and was compensated a longer experiment time to record more scans. The spectrum of eIF4A changes considerably upon binding, with many chemical shift changes, and the broader lines expected for the higher molecular weight and longer rotational correlation time. Despite the broader lines, the spectrum is well-dispersed and looks promising for further NMR studies. Moreover, the lack of precipitation in the 500μM NMR sample suggests that even higher concentrations might be achievable.

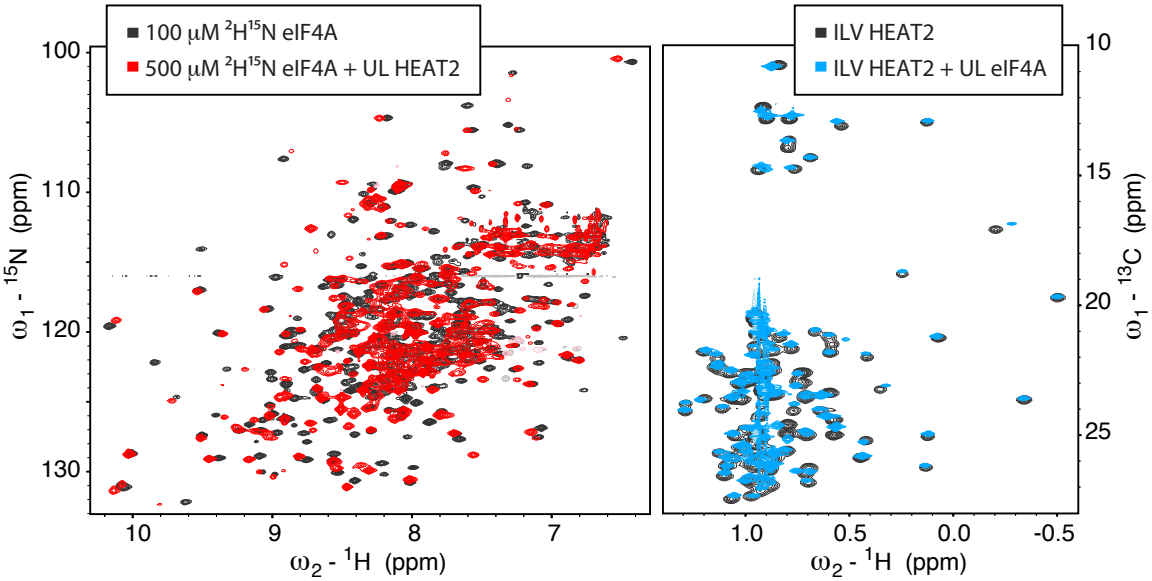

Figure 5.2: The unlabeled eIF4G HEAT2 domain causes significant shifts in the ^{15}N HSQC of eIF4A, as well as making eIF4A much less prone to aggregation (left panel). Likewise, eIF4A causes significant shifts in spectra of HEAT2. The right panel shows a ^{13}C HSQC of ILV-labeled HEAT2, in the presence (blue) and absence (black) of eIF4A.

The right panel of Figure 5.2 shows spectral changes from the other side of the complex, and with another labeling scheme. In these spectra, HEAT2 is ILV-labeled. In other words, it is fully deuterated, with ^{14}N and ^{12}C, except for ^{1}H and ^{13}C at the methyl positions of all isoleucine, leucine, and valine residues. The ^{13}C HSQC of ILV-HEAT2 alone is shown in black contours, and the ^{13}C HSQC of ILV-HEAT2 in complex with unlabeled eIF4A

is shown in cyan contours. Many of the methyl resonances are unperturbed, but several notable peaks shift. In particular, the δ methyl resonance for I1334 and both methyl peaks for V1337 shift considerably.

5.1.2 Backbone assignments

In an effort to assign the backbone chemical shifts of eIF4A in complex with the HEAT2 domain, we recorded several of the standard backbone experiments, including the HNCA, HNCO, HNCACO, and HN(CA)CB. All spectra were recorded using TROSY and nonuniform sampling, and processed using Forward Maximum entropy reconstruction [26]. These spectra all exhibit significant dynamic range, due to the variety of relaxation rates in different parts of the complex. Some flexible loops give very strong signals, while the most rigid parts of the protein have acceptable signal/noise, and many parts of the protein appear to have line-broadening caused by conformational exchange. This combination makes it challenging to predict the ideal number of scans for the 3-dimensional experiments based on the first increments; ultimately the spectra are missing too many of the weak peaks for a confident assignment. Some parts of the spectra are quite useful, however, and we were able to assign many of the well-dispersed peaks around the periphery of the TROSY-HSQC. Figure 5.3 shows a series of strips in the HNCA spectrum that we were able to assign with confidence. Unfortunately, many of the peaks for residues in the interface appear to be missing, probably due to conformational exchange, which may complicate further structural studies.

5.1.3 Structural restraints

We can see that the interaction surface between these two proteins must be quite large, giving us hope that we might be able to measure NOEs between residues on opposite sides

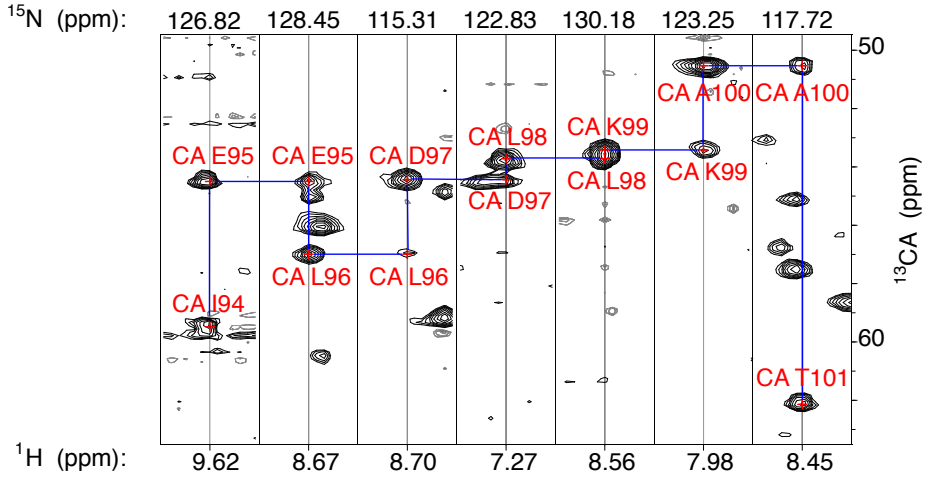

Figure 5.3: HNCA assignments of $^{2}H^{15}N^{13}C$ eIF4A in complex with HEAT2. Representative segment far from the binding interface.

of the interface. Several difficulties arise, however. First, extensive deuteration is necessary to prevent drastic relaxation losses in a 70 kDa complex. Second, we need a strategy for limiting the information content to a manageable amount: we would like to be able to distinguish intermolecular NOEs from intramolecular NOEs. We hoped to achieve both of these goals by perdeuterating and ^{15}N-labeling one protein, and ILV-labeling the other binding partner. This way, we can ignore amide-amide and methyl-methyl intramolecular NOEs, and focus on the amide-methyl and methyl-amide signals of interest. The observed methyl shifts in Figure 5.2 suggest that we may be able to measure methyl-methyl or methyl-amide NOEs at the interface, despite the predominantly polar interface. Based on our preliminary models, there seems be a lack of ILV residues near the HEAT2-binding interface in eIF4A. HEAT2 has a few plausible candidates, so we first tried ILV methyl-labeling of HEAT2 and ^{15}N labeling of eIF4A. The predominantly α-helical nature of both proteins leads us to be concerned that the amide protons nearest to the interface may still be buried deep within helices, too far for NOEs. On the other hand, the interhelical loops

appear to play a significant role in binding, so amide labeling is worth trying. We recorded a time-shared 3D HSQC-NOESY in an effort to measure the intermolecular NOEs with ^{13}C-dispersion and ^{15}N-dispersion simultaneously [16], but we were unable to observe any intermolecular NOEs.

Our model for HEAT2 binding to eIF4A in the open form (Section 5.3.5 and Figure 5.14), suggests that I1332, I1334, and V1337 in HEAT2 are likely to be near the aromatic sidechain of F179 in eIF4A. This is consistent with the strong amide and methyl chemical shift changes that we see in this region, and leads us to hope that we may measure NOEs between at least one of these methyl groups and the F179 sidechain, by recording an HMQC-NOESY spectrum on a sample with ILV-HEAT2 and ^{2}H eIF4A with selective protonation of phenylalanine sidechains.

5.2 Measuring eIF4A domain separation with FRET

Another method which can answer some of our questions about whether eIF4A is in an open, closed, or intermediate state, while accommodating dynamics, is measuring single-molecule Fluorescence Resonance Energy Transfer (FRET). FRET enables us to measure distances between points on opposite domains of eIF4A under a variety of conditions. Using single-molecule FRET, rather than the bulk technique, gives us a better understanding of the variety of states that occur under particular conditions [58]. We prepared a panel of fourteen double-cysteine mutants of eIF4A, with sites for attaching the dyes distributed over the surface of the domains. Figure 5.4 shows a cartoon illustration of how we expect to use FRET to evaluate our current models of eIF4A interdomain conformations. All the models depicted in Figure 5.4 are based on static crystal structures [2, 8, 9, 52], yet we expect that the true behavior in solution is more dynamic. Our initial model for eIF4A structure and function suggests that the activity of eIF4A hinges on its ability to cycle through a variety

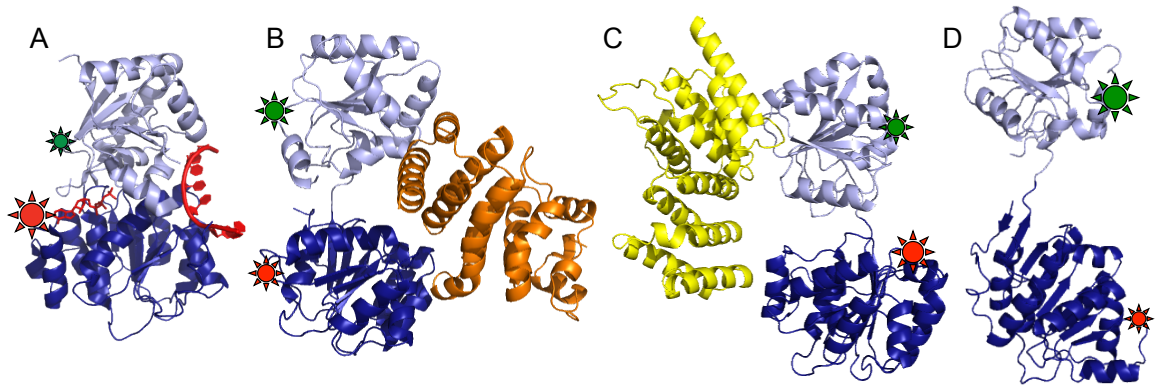

Figure 5.4: Cartoon illustration of how FRET results will help us evaluate our models of eIF4A conformations under a variety of conditions in solution, ranging from a closed conformation (A), through a variety of partially open states (B,C), to a wide-open state. In each panel, eIF4A-NTD is colored dark blue and eIF4A-CTD is colored light blue. HEAT2 is colored orange and HEAT1 is colored yellow. Green and red bursts in each panel represent the donor and acceptor fluorophores, sized according to various levels of FRET efficiency based on the distance between them.

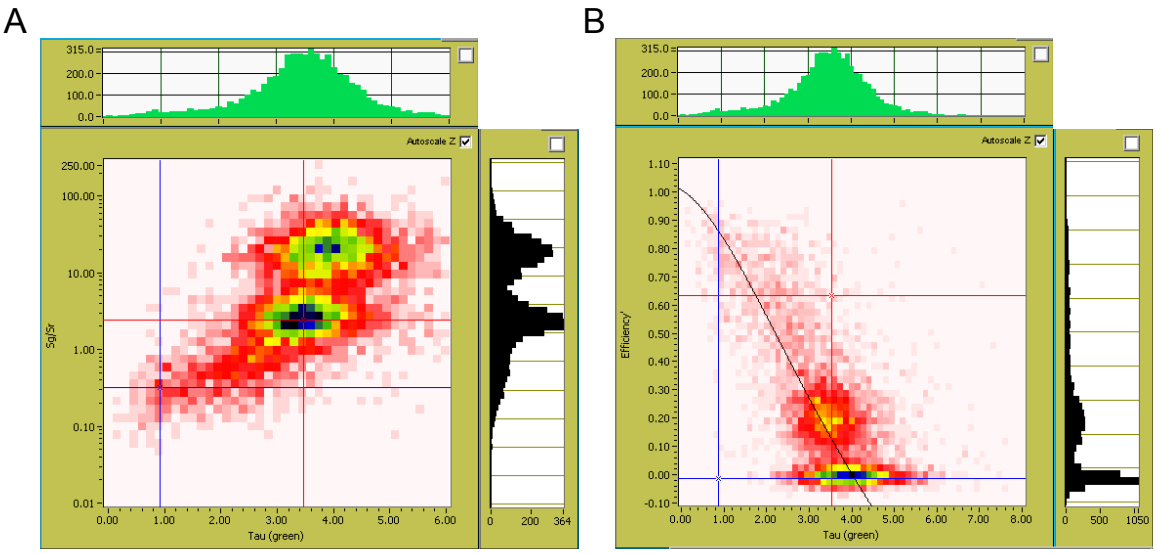

Figure 5.5: A. 2D histogram correlating the signal intensity ratio on the green and red channels (Sg/Sr) to green fluorescence lifetime (tau) for free eIF4A with fluorophores attached at residues 68 and 287. B. FRET efficiency vs. donor fluorescence lifetime for the same sample. The large population with high Sg/Sr and high green fluorescence lifetime, and Efficiency=0 is free dye or protein labeled with only the donor fluorophore. Most doubly labeled molecules have FRET efficiency between 0.1 and 0.3, indicative of a wide open conformation, but some are more compact. Figure prepared by Thomas-Otavio Peulen.

of open and closed states according to conditions (Figure 1.4), and single-molecule FRET will enable us to observe these different populations.

So far, our FRET experiments have been hindered by the surprisingly low expression yields and solubility of the mutants, but we have managed to obtain some initial results for a single mutant eIF4A without a binding partner. We observe that free eIF4A mostly occupies a fully-extended conformation in solution, much like the crystal structure of yeast eIF4A [8]. Unsurprisingly, we also see evidence that it samples more closed states occasionally. Figure 5.5 shows histograms of eIF4A conformations, plotted according to the signal intensity ratio on the green and red channels (Sg/Sr) and FRET efficiency. The dominant state is wide open, with 60Å between fluorophores (center peak with Sg/Sr between 1.0 and 10.0 in panel A and 0.1-0.3 Efficiency in panel B). There is also evidence for dynamic interconversion with a more closed conformation, with 45Å between fluorophores (smear with Sg/Sr between 0.1 and 1.0 (A), and FRET efficiency between 0.4 and 0.9 (B)). In this case, challenges in sample preparation prevented us from ensuring that each protein was labeled with a single fluorophore of each color; proteins labeled with only green dye appear in the top peak in panel A, with high Sg/Sr, and in the sharp peak at the bottom of panel B, with 0 FRET efficiency.

5.3 Combining pairwise interactions into a large model

In this section, we expand upon the pairwise interactions that we mapped in Chapter 4, combining these interactions to form a model of the entire eIF4A/eIF4G/eIF4H complex, following the starting model presented in Figure 1.6. We begin modeling using the closed helicase conformation, shown in the right panel of Figure 1.4. Next, we address questions relating to the dynamics of the complex, with competitive binding interactions throughout the assembly, and how the dynamics relate to helicase function. Finally, we present another

model for the open form of the helicase, which most closely resembles what we expect to find when we solve the NMR structure of full-length eIF4A bound to the HEAT2 domain of eIF4G.

5.3.1 Modeling the complex onto a closed helicase structure

In Chapter 4, we used chemical shift perturbations to map interactions involving eIF4A-NTD and the HEAT2 domain of eIF4G. We combine this information with binding sites for HEAT1, HEAT2, eIF4H, and the HEAT1-HEAT2 linker, all mapped on the surface of eIF4A-CTD (shown in Figure 5.6A). Although binding between eIF4A-NTD and the HEAT1 domain of eIF4G was too weak for us to map using chemical shift perturbations, the binding interaction was confirmed by mutagenesis and surface plasmon resonance experiments. Figure 5.6B shows data from surface plasmon resonance (SPR) experiments of wildtype and mutant HEAT1 binding to eIF4A. The F978A mutation in HEAT1 dramatically reduces binding to full-length eIF4A (red line), although the mutation has little effect on binding to eIF4A-CTD (data not shown). Although we cannot map the interface by NMR, we can conclude that F978 is in or near the eiF4A-NTD binding site on the HEAT1 domain. These interfaces are all compatible with our original model of the domain organization of eIF4G, and how a closed form of eIF4A can fit sandwiched between HEAT1 and HEAT2 of eIF4G, along with eIF4H (Figure 1.6).

Figure 5.7 shows the model that results after combining the closed form model of the eIF4A domains with the predicted arrangement of the HEAT domains of eIF4G and eIF4H. The relative positions of the domains were adjusted to conform to the binding interfaces measured by NMR. Domains with known structures or homology to known structures were modeled as ellipsoids of appropriate relative sizes. The model omits any indication of how eIF4H might interact with the N-terminal domain of eIF4A because we were unable to map

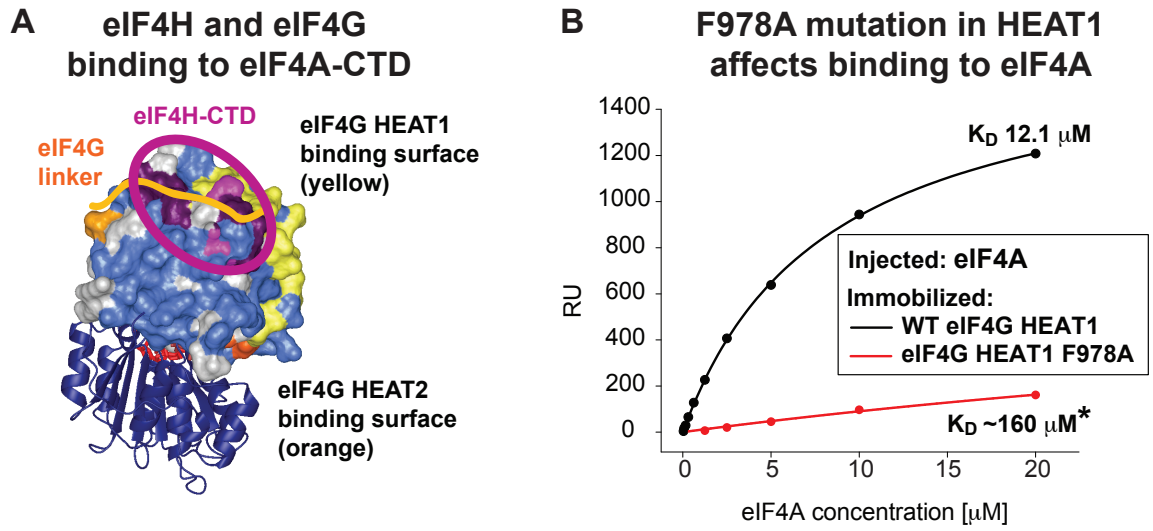

Figure 5.6: Additional binding interfaces among eIF4A, eIF4G, and eIF4H domains. (A) eIF4G and eIF4H binding surfaces mapped on eIF4A-CTD. The HEAT1 binding interface is shown in yellow, the HEAT2 interface is shown in dark orange. The binding surfaces of the C-terminal part of eIF4H overlaps with the binding surface of the linker between HEAT1 and HEAT2 of eIF4G. Residues affected only by the linker are shown in light orange. Residues affected only by eIF4H are shown as light purple. Dark purple residues are affected by both eIF4H and the eif4G linker. (B) Graphs of SPR data showing binding of full-length eIF4A to immobilized WT eIF4G HEAT1 (black) and HEAT1 F978A (red). The K_D value for the mutant is marked with a star because it should be considered only an estimate, since concentrations higher than the K_D could not be reached in the titration due to limited solubility (figures adapted from Marintchev et al. [42]).

the binding interfaces, although there is ample evidence that the interaction does occur in the presence of ATP [42, 50], and the model is based on a nucleotide-bound form of eIF4A.

This model contains three previously identified RNA binding sites: in the interdomain cleft of eIF4A, the RRM of eIF4H, and on the HEAT1 domain of eIF4G [40]. Tracing a possible path for the RNA through these domains leads to the particular appeal of this model. In crystal structures of eIF4A homologs bound to RNA, the 5' end of the RNA contacts the C-terminal domain, and the 3' end contacts the N-terminal domain, so that a helicase scanning from the 5' end toward the 3' end of a messenger RNA would scan with the N-terminal domain in front [2]. The eIF4H RRM domain is in position to capture the

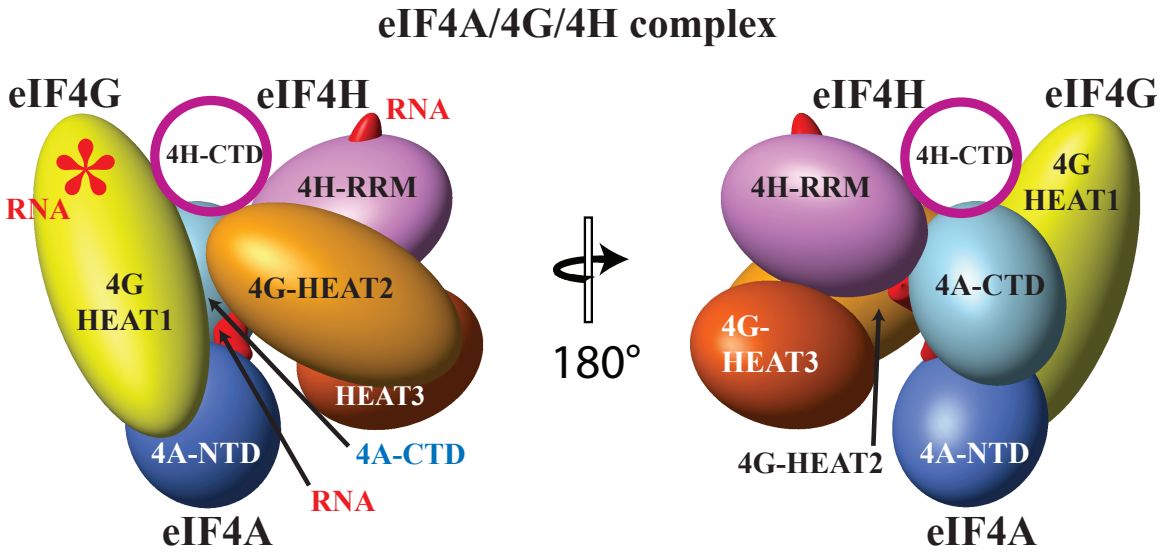

Figure 5.7: Organization of the eIF4A/eIF4G/eIF4H helicase complex (adapted from Marintchev et al. [42]). This model of the ternary complex of eIF4A, eIF4G, and eIF4H illustrates how the domains of these proteins could be arranged in agreement with chemical shift mapping and mutagenesis data. Previously identified RNA binding sites on eIF4A, eIF4G HEAT1, and eIF4H are indicated in red.

RNA immediately after unwinding by the helicase. Unlike the eIF4H RNA binding site, the binding site on HEAT1 does not appear to be in the direct path of the RNA strand bound in the eIF4A cleft, so the RNA might loop back into this site somehow. Finally, we should recall that on a messenger RNA, this whole complex might still be attached to the 5' end of the mRNA via the m^7GTP cap, eIF4E, and the eIF4E binding site in the N-terminal third of eIF4G. Figure 5.8 illustrates how the mRNA might wind through the complex, from the 5' cap to HEAT1, then looping around to eIF4H and finally the helicase. This figure also illustrates how if a sufficiently large loop of RNA develops after unwinding by the helicase, it may accommodate the small ribosomal subunit.

Prior models of scanning placed the whole eIF4F complex behind a scanning small ribosomal subunit on the mRNA, due to the attachment of eIF4E to the 5' cap. It was unclear how the eIF4A helicase could unwind secondary structure to facilitate subunit scanning,

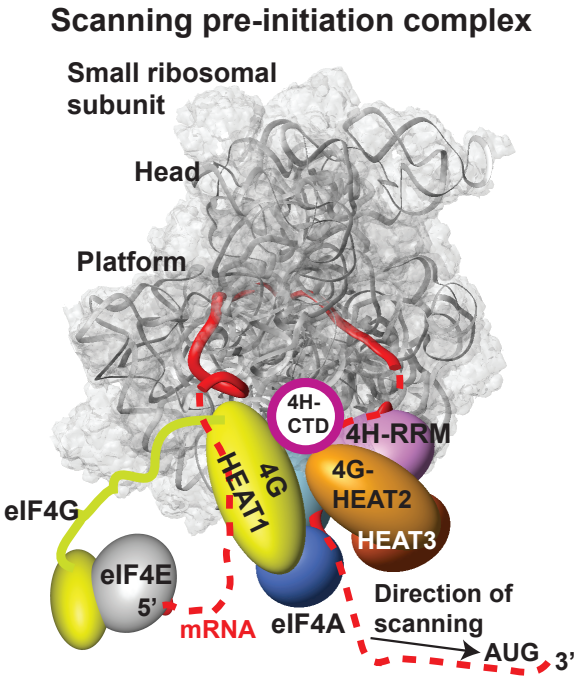

Figure 5.8: Model of the scanning pre-initiation complex (adapted from Marintchev et al. [42]). The model shown in Figure 5.7 is expanded to show how the small ribosomal subunit might fit (gray semi-transparent surface, with the rRNA backbone shown as ribbon). The ribosome with bound mRNA (red solid ribbon) are taken from PDB 1JGP [71]. eIF4E and a larger section of eIF4G are also shown, along with a potential path for the mRNA (dashed red line).

while following behind the subunit. This new model is especially appealing because it illustrates how eIF4F can remain attached to the 5' cap of the mRNA, yet unwind secondary structure ahead of a scanning small ribosomal subunit.

5.3.2 Competitive binding in the translation initiation complex

We have mapped binding sites and explored the interactions of both eIF4A domains with ATP, RNA, eIF4H, the HEAT1 domain of eIF4G, the HEAT2 domain of eIF4G, and the linker between HEAT1 and HEAT2. Among these interactions, we have observed some that appear to share an interface, and therefore can be expected to compete for binding to eIF4A. Section 4.4 addressed how HEAT2 and RNA share a binding surface on eIF4A-NTD, and

the binding surface described for HEAT2 on eIF4A-CTD also corresponds to the expected RNA binding site [42]. HEAT2 indeed competes with RNA for binding to full-length eIF4A (Figure 4.14b).

Some of these interactions have been shown to be cooperative: it is well-known that ATP stimulates RNA binding to eIF4A and vice-versa [39], and eIF4H has been shown to bind cooperatively with ATP [42]. HEAT1 has previously been proposed to enhance eIF4A helicase activity by stabilizing a more closed form, which would more readily bind to ATP [45]. Consistent with this speculation, HEAT1 binds to eIF4A more tightly in the presence of ATP, as shown by surface plasmon resonance in Figure 5.9A. Since ATP and RNA binding to eIF4A are coupled, we can speculate that the HEAT2 domain interferes with binding to ATP, and ATP interferes with binding to HEAT2 as well. Indeed, HEAT2 binds more tightly to eIF4A in the absence of ATP than in the presence of ATP, as shown in Figure 5.9B. Figure 5.10 summarizes the effects each cofactor has on the nucleotide affinity of eIF4A. HEAT1, eIF4H, and RNA stimulate ATP binding, and therefore we assume they promote or have affinity for the closed state of the helicase. HEAT2, on the other hand, interferes with ATP binding, so we claim that it promotes an open state of the helicase, which has decreased nucleotide affinity due to the separation of ATP-binding motifs on opposite domains of eIF4A.

If the HEAT1 domain induces a closed state of the helicase, while the HEAT2 domain binds preferentially to the open state, then the two domains should compete for binding, despite having non-overlapping binding sites on eIF4A. To test this hypothesis, we used isothermal titration calorimetry (ITC) to measure the affinity of the HEAT2 domain for eIF4A, in the presence and absence of HEAT1. Figure 5.11A shows the ITC binding response and curve fit for injecting HEAT2 into eIF4A, resulting in a 0.9μM dissociation constant for the HEAT2:eIF4A complex. Figure 5.11B shows the measured binding of

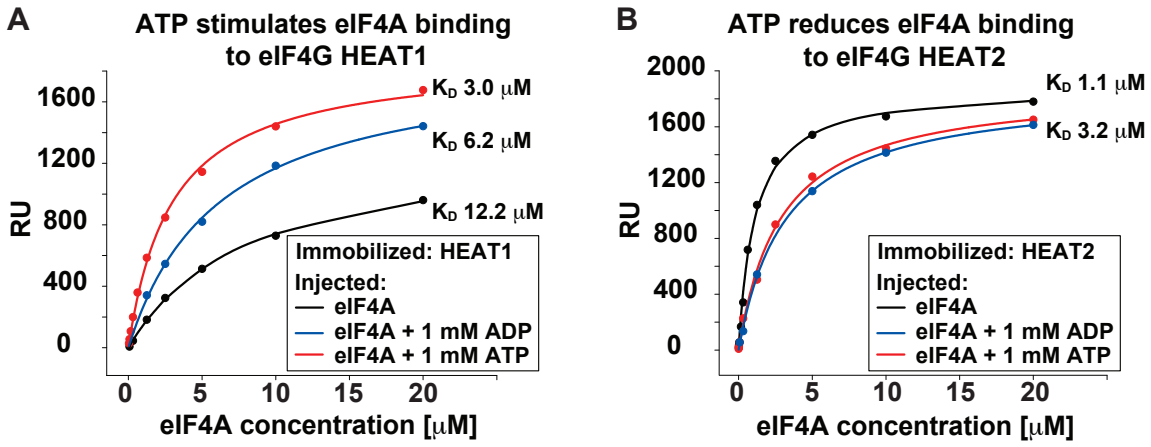

Figure 5.9: Surface plasmon resonance graphs of eIF4A binding to immobilized eIF4G HEAT1 (A) or eIF4G HEAT2 (B) in the absence of nucleotide (black) and in the presence of 1 mM ATP (red) or 1 mM ADP (blue) (adapted from Marintchev et al. [42])

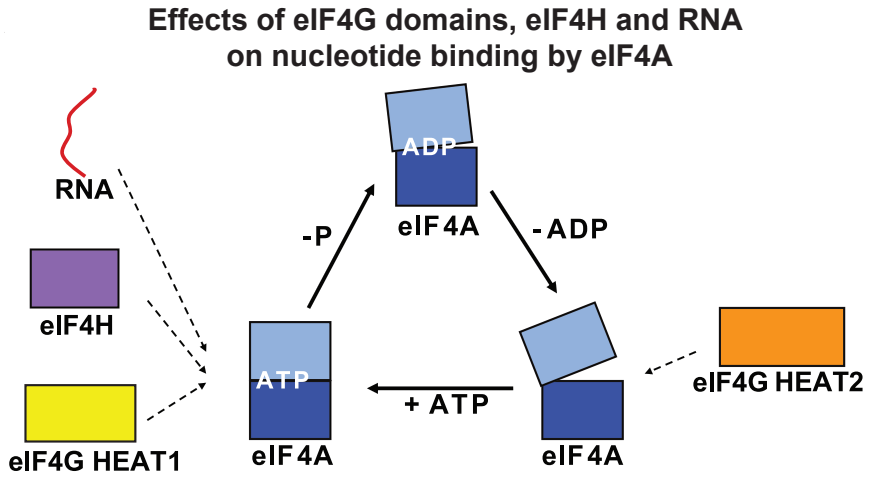

Figure 5.10: Summary of the effects of eIF4G, eIF4H and RNA on the affinity of eIF4A for nucleotides. RNA, eIF4H, and eIF4G HEAT1 (left) favor the closed, ATP-bound conformation, whereas HEAT2 (right) promotes the open nucleotide-free conformation of eIF4A (adapted from Marintchev et al. [42])

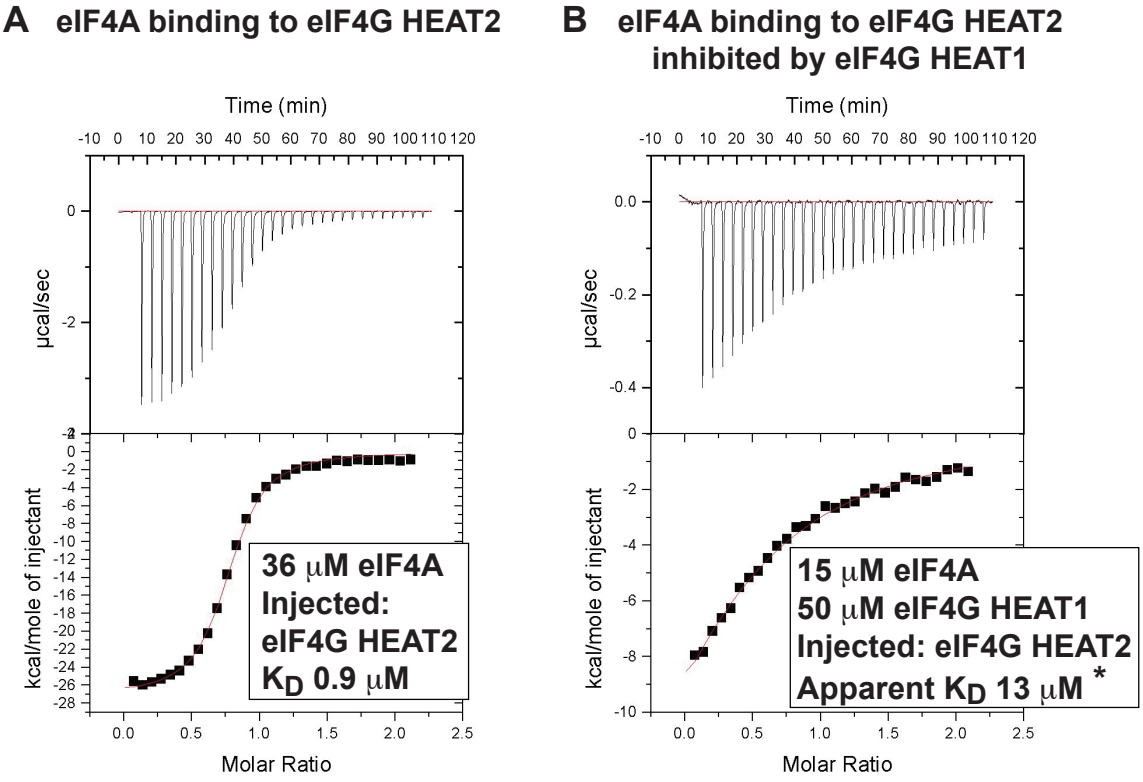

Figure 5.11: Isothermal Titration Calorimetry (ITC) graphs of eIF4G HEAT2 binding to eIF4A in the absence (A) and presence (B) of eIF4G HEAT1. Note that in the experiment shown in panel (B), the concentration of eIF4G HEAT1 (50 μM) is not saturating. Therefore, a fraction of eIF4A is not bound to HEAT1 and the calculated apparent K_D for the interaction between eIF4A and eIF4G HEAT2 in the presence of HEAT1 (marked with an asterisk) should be considered a lower limit of the actual K_D (adapted from Marintchev et al. [42]).

HEAT2 to eIF4A in the presence of excess HEAT1 domain, resulting in a dissociation constant of 13μM. Although the HEAT1 was not present in a high enough concentration to saturate eIF4A in the absence of ATP, it decreased the affinity of eIF4A for HEAT2 by more than ten-fold. Saturating levels of HEAT1 should have a greater effect.

The HEAT1 and HEAT2 domains of eIF4G bind anticooperatively to eIF4A, probably because HEAT1 can bind to both domains of eIF4A only in a relatively close conformation, while HEAT2 can bind maximally to eIF4A only in a relatively open conformation. Therefore, only one eIF4G HEAT domain at a time can bind both eIF4A domains with high affinity, while the other HEAT domain must either be bound to only one of the eIF4A domains or its binding to the second domain is suboptimal. We speculate that HEAT2 might compete with eIF4H binding in the presence of ATP by a similar mechanism. Moreover, it appears that eIF4H and the linker between HEAT1 and HEAT2 might compete directly for binding to the same surface on eIF4A-CTD [42]. All of these anticooperative binding events raise some doubts about whether the ternary eIF4G/eIF4H/eIF4A complex presented above in Figure 5.7 can really exist. We explore this question in more detail in the next section.

5.3.3 eIF4A, eIF4G, and eIF4H form a stable complex in the presence of ATP

We used size-exclusion chromatography on mixtures of eIF4A, eIF4H, and eIF4G to explore whether these proteins form a stable ternary complex. The eIF4G construct used, "4G-Se", corresponds to the C-terminal two-thirds of eIF4G and contains all known eIF4A-binding sites. We first confirmed that eIF4A and eIF4H form a stable binary complex in the presence of ATP (pink line in Figure 5.12A). Individually, eIF4A and eIF4H each elute starting in fraction 17 (not shown), but when these proteins are combined in the presence of ATP, SDS-PAGE analysis of the fractions shows that they form a larger complex that elutes

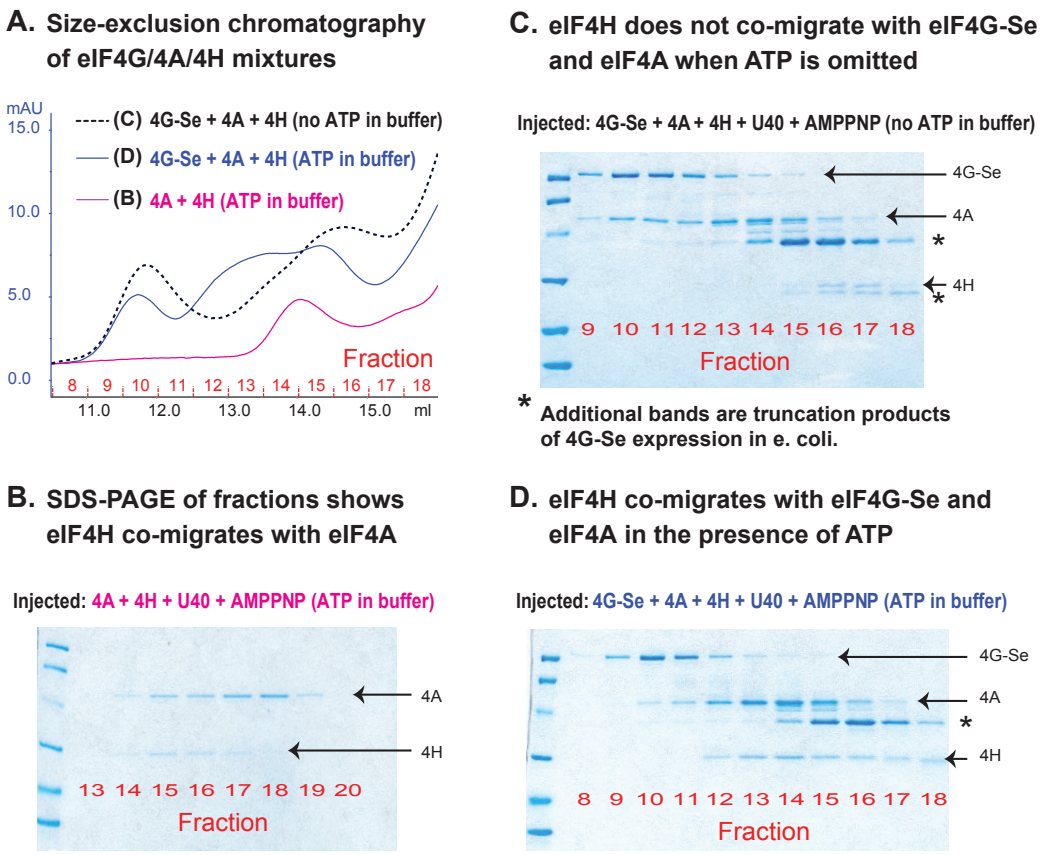

Figure 5.12: Size-exclusion chromatography of eIF4G/4A/4H ternary complex (adapted from Marintchev et al. [42]).
A. Chromatograms from analytical gel filtration of a mixture of 4G-Se, eIF4A, and eIF4H, run with and without ATP in the buffer (blue line and dashed black line, respectively), and of a mixture of eIF4A and eIF4H, run with ATP in the buffer (pink line). Absorption at 280 nm is plotted versus elution time and fraction number for the three mixtures. The 4G-Se construct corresponds to the C-terminal two-thirds of eIF4G.
B. SDS-PAGE analysis of fractions collected from separation of the mixture of eIF4A and eIF4H. The two proteins begin to elute in fraction 14, earlier than they elute when run separately (fraction 17; data not shown), indicating that they form a complex.
C. SDS-PAGE analysis of fractions collected from separation of the 4G-Se/eIF4A/eIF4H mixture, without ATP in the running buffer. Bands of 4G-Se, eIF4A, and eIF4H are indicated. Remaining bands are truncation products of 4G-Se, marked with asterisks. eIF4A elutes as early as 4G-Se, beginning in fraction 9, indicating binary complex formation. In the absence of ATP, eIF4H does not join the complex: it begins to elute only in fraction 15.
D. SDS-PAGE analysis of fractions collected from separation of the 4G-Se/eIF4A/eIF4H mixture, with ATP in the running buffer. Again, eIF4A elutes early with 4G-Se, indicating tight complex formation. Now, eIF4H elutes earlier than in the binary eIF4A/eIF4H complex, beginning in fraction 12.

earlier, starting in fraction 14 (Figure 5.12B).

When all three proteins, eIF4A, eIF4H, and 4G-Se, are analyzed together, but ATP is omitted from the running buffer, we observe a complex of eIF4A and 4G-Se eluting together beginning in fraction 9, while eIF4H elutes later, in fraction 15 (Figure 5.12C). On the other hand, including ATP in the running buffer brings eIF4H into the ternary complex with eIF4A and 4G-Se: in this case, the eIF4H elutes as early as fraction 12, indicating that it is part of a larger complex, compared to the eIF4A-eIF4H complex which elutes starting in fraction 14 (Figure 5.12D). Although eIF4H does not bind tightly enough to eIF4A in this context to elute from the size-exclusion column as early as fraction 9, this elution profile is evidence that the ternary complex does form.

5.3.4 Dynamics within the translation initiation complex

We have shown that eIF4A, eIF4H, and eIF4G form a stable complex in the presence of ATP and RNA, despite several cases of anticooperative binding among the domains involved. Of the many small interdomain interfaces we have described which hold this complex together, not all can be satisfied simultaneously. Therefore, these proteins undergo dynamic rearrangements within the context of the larger complex, especially as conditions change. In particular, we believe that these rearrangements are critical for the nucleotide binding, hydrolysis, and release processes.

Figure 5.13 illustrates the interplay between the dynamics of the eIF4A/eIF4G interactions and the ATP binding and hydrolysis cycle. In the absence of nucleotide, eIF4A binds to eIF4G HEAT2 more tightly than to HEAT1, resulting in the open state of eIF4A shown in the right panel of Figure 5.13. Coupled binding of ATP and RNA should displace HEAT2 and bring eIF4A-NTD and eIF4A-CTD together toward the closed state shown in the left panel. Hydrolysis of ATP and release of ADP should shift the helicase back to the open

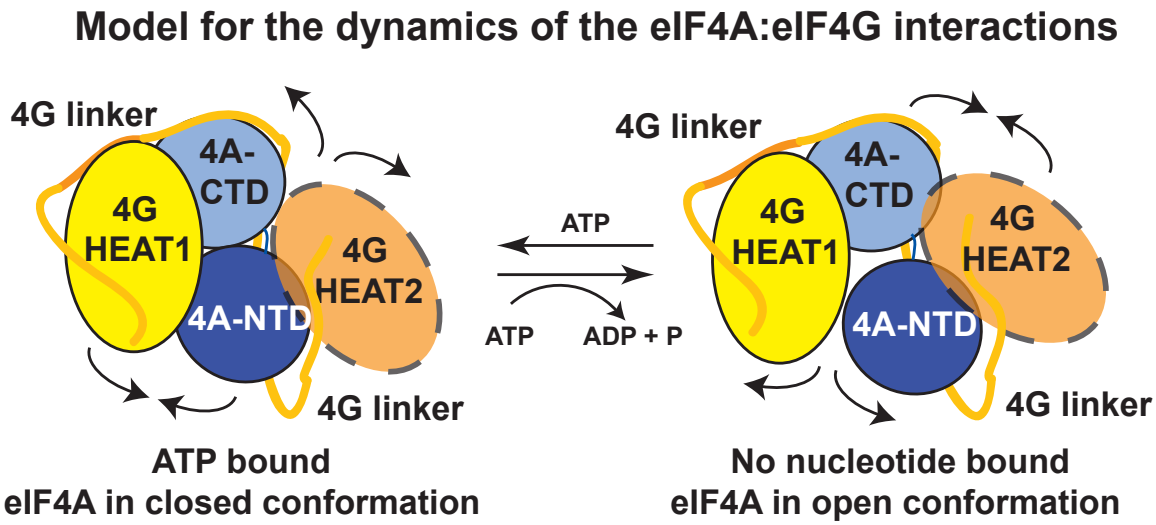

Figure 5.13: Model for the dynamics of the eIF4A/eIF4G interactions. eIF4G HEAT1 stimulates ATP binding and the helicase activity of eIF4A by simultaneous binding to both eIF4A domains in the closed, ATP-bound conformation (left). eIF4G HEAT2 promotes the nucleotide-free state by binding to both eIF4A domains in an open conformation (right). The arrows indicate directions of rearrangements during the ATP hydrolysis and nucleotide exchange cycle. ATP (not shown) binds at the interface between the two eIF4A domains. (adapted from Marintchev et al. [42])

form on the right.

5.3.5 Modeling the complex onto an open helicase structure

The model discussed above contains eIF4A in a closed conformation, which we believe is a state that the helicase occupies during the ATP-hydrolysis process, when it is bound tightly to ATP and to RNA. We also described more open states that the helicase must cycle through, as it releases ADP and binds again to ATP, and we presented evidence that the HEAT2 domain stabilizes at least one of these more open states. Since that time, two crystal structures have been solved of human eIF4A in complex with Pdcd4 [9, 37], which contains two domains that are homologous to the HEAT2 domain of eIF4G. Figure 1.5 shows the alignment of HEAT2 with both MA3 domains of Pdcd4. These structures confirm previous mutagenesis data that suggest that Pdcd4 binds to eIF4A through residues that

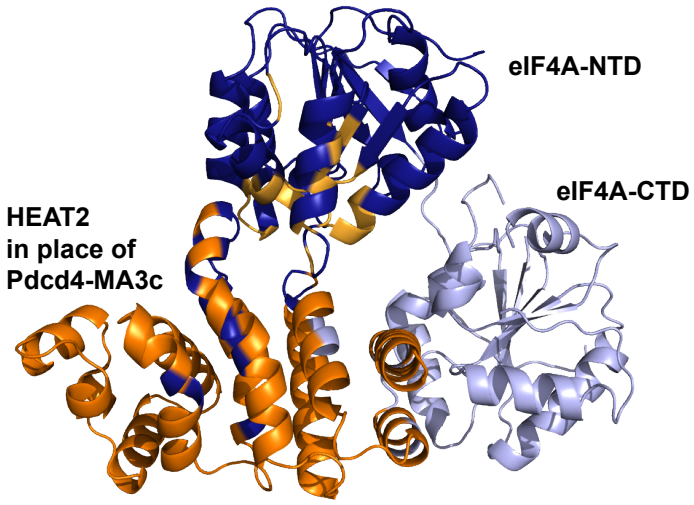

Figure 5.14: Model of eIF4G-HEAT2:eIF4A complex based on the crystal structure of homologous PDCD4 HEAT domains bound to eIF4A. Residues in HEAT2 with significant chemical shift changes in the titration with eIF4A-NTD are colored navy, and those with significant changes caused by eIF4A-CTD are colored light blue. Residues in eIF4A-NTD that are strongly affected by HEAT2 binding are colored gold.

have counterparts in HEAT2 [70], as well as our NMR data that show that HEAT2 and PDCD4 both bind to the RNA-binding region of eIF4A-NTD (Figures 4.13 and 4.14.

We can exploit this homology to model HEAT2 in the place of one of the Pdcd4 domains in one of these crystal structures. The two structures are quite similar, but Loh et al. crystallized mouse Pdcd4 with human eIF4A [37], whereas Chang et al. crystallized both human constructs, and were able to obtain higher resolution, so we chose this structure (PDB ID 2ZU6) [9]. We used Pymol [51] to align HEAT2 to one of the MA3m domains in the asymmetric unit, producing the model in Figure 4.4. The same model is shown from another perspective in Figure 5.14. Here we also indicate the residues identified as involved in binding in the pairwise interaction studies described in the previous chapter: HEAT2 residues with strong chemical shift perturbations in titrations with eIF4A-NTD and eIF4A-CTD are highlighted in dark and light blue, respectively, while eIF4A-NTD residues with significant shifts upon binding HEAT2 are painted gold. This model appears

to be consistent with the chemical shift perturbation data from Chapter 4. The separation between the two domains of eIF4A is substantial in this model.

6

Conclusions and future directions

6.1 Concluding remarks

In this thesis, we have presented the NMR solution structure of the HEAT2 domain of eIF4G. The structure differs slightly from the one which was previously solved by X-ray crystallography, by virtue of being in a more native-like environment, without crystal packing artifacts.

We have used NMR chemical shift mapping to explore binding interfaces involving the HEAT2 domain of eIF4G and the N-terminal domain of eIF4A, with other binding part-

ners. Using the interfaces determined for eIF4G HEAT2 binding to eIF4A-NTD along with distance restraints obtained using site-directed spin labeling and paramagnetic relaxation enhancement, we obtain a structure of the two-domain complex, and are able to speculate about how it might fit into a structure involving the full-length eIF4A helicase, and how the binding mode of the HEAT2 domain might compare to those of the structurally similar Pdcd4 MA3 domains.

We also present work on larger and more functional complexes involving eIF4A, eIF4G, and eIF4H. This includes work towards an NMR structure of the 70kDa complex of full-length eIF4A with the HEAT2 domain of eIF4G. We also explore the possibility of using single-molecule FRET techniques to measure the separation of the domains in eIF4A under different conditions, as it relates to helicase function. Finally, we combine all the pairwise interactions among domains measured by chemical shift mapping and other techniques to construct a structural and functional model of the complex and how it cycles through open and closed states to unwind RNA.

6.2 Further structural characterization of the eIF4A:HEAT2 complex by NMR

Work is ongoing to complete the NMR backbone assignments of eIF4A and HEAT2 in the complex. We have prepared a panel of single-cysteine mutants of HEAT2 for site-directed spin-labeling, and are in the process of measuring paramagnetic relaxation enhancements on these samples. We may also attach paramagnetic spin-labels to eIF4A to measure the effect on HEAT2, but the low expression yields and low solubility of eIF4A single-cysteine mutants give some cause for concern. The lower solubility may be somewhat mitigated by the ability to measure the paramagnetic effects on the methyl ^{13}C HSQC with ILV-labeled HEAT2. We are now using not only the N-terminal truncation mutant of eIF4A, but also

the truncated form of the HEAT2 domain, described in Section 3.2.3, which we can expect not to cause the T1 noise that mars the spectrum shown in the right panel of Figure 5.2, at 0.9 ppm.

6.3 Continuing FRET studies of eIF4A complexes

We may have an opportunity to continue the FRET studies on eIF4A with more mutants and under more conditions if we can solve some of the difficulties that we have encountered in expression and solubility. The single-cysteine mutants we have used have mutations not only to the native cysteine residues, but also to several aromatic residues near the introduced cysteine residues, to prevent interference with the fluorescence of the attached fluorophores. Compromising on this point may be necessary to obtain data, although it may yield less reliable data than we could theoretically obtain with the ideal samples that we are unable to prepare.

Such a compromise may not be necessary, however. The sample quantity and solubility requirements for the single-molecule FRET experiments are quite low; the problem arises during the labeling process. We would like to prepare samples which uniformly have two fluorophores attached, one of each color. This can be achieved by adding the two dyes sequentially, with ion exchange chromatography after each step, to purify only protein with one fluorophore attached after the first labeling step, and then only protein with two fluorophores attached after the second step. Unfortunately, most of the protein precipitates under the low-salt conditions required for binding eIF4A to the column.

A possible solution to this problem would be to bind the eIF4A double-cysteine mutants to a cysteine-free mutant of HEAT2 prior to the labeling reactions. Since this complex is typically much more soluble than free eIF4A, and is especially soluble under low-salt conditions, it may enable the entire fluorophore labeling process. If necessary, the HEAT2 can

be removed afterwards by size-exclusion chromatography or nickel-NTA chromatography under high salt conditions.

6.4 Biological importance

Our work on the interaction of HEAT2 with eIF4A has given us a better understanding of how it might exert a mild inhibitory effect on eIF4A, through direct competition for binding to RNA, and allosteric competition for ATP binding. When supplied in *trans*, it additionally reduces activity by interfering allosterically with eIF4A binding to HEAT1. We speculate that the inhibitory effect of HEAT2 is considerably less than that of Pdcd4 for several reasons. First, it improves overall binding of eIF4G to eIF4A, thus making activation by the HEAT1 domain more available. Second, it appears that HEAT2 might bind in a slightly different mode than Pdcd4, which may influence its ability to hold the two domains apart, in an inactive state. Moreover, the presence of the HEAT2 domain also serves to compete away the inactivating Pdcd4 domains. The HEAT2 domain appears to contribute to maintenance of a delicate balance between activation and repression of translation.

The new model for how eIF4A facilitates preinitiation complex scanning through the 5'-UTR of a messenger RNA is a clear breakthrough in our understanding of the mechanism of cap-dependent translation initiation, although many of the critical details have yet to be verified. Our previous understanding placed the unwinding complex of eIF4G and eIF4A upstream of the scanning ribosome, because of the attachment of eIF4G and eIF4A to the 5' cap through eIF4E. It was therefore unclear how eIF4A could facilitate scanning by unwinding secondary structure in RNA that has already passed through the scanning preinitiation complex. This new, more compelling, model places the helicase functionality on the mRNA ahead of the scanning 40S ribosome, while still permitting interactions with the 5' cap.

References

[1] MestRe Nova. URL `http://www.mestrelab.com`.

[2] Christian B F Andersen, Lionel Ballut, Jesper S Johansen, Hala Chamieh, Klaus H Nielsen, Cristiano L P Oliveira, Jan Skov Pedersen, Bertrand Séraphin, Hervé Le Hir, and Gregers Rom Andersen. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*, 313(5795):1968–1972, September 2006.

[3] Sharon J Archer, Mitsuhiko Ikura, Dennis A Torchia, and Ad Bax. An alternative 3D NMR technique for correlating backbone 15N with side chain H resonances in larger proteins. *Journal of Magnetic Resonance*, 95(3):636–641, December 1991.

[4] John L Battiste and Gerhard Wagner. Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data†. *Biochemistry*, 39(18):5355–5365, May 2000.

[5] Lluís Bellsolell, Park F Cho-Park, Francis Poulin, Nahum Sonenberg, and Stephen K Burley. Two structurally atypical HEAT domains in the C-terminal portion of human eIF4G support binding to eIF4A and Mnk1. *Structure*, 14(5):913–923, May 2006.

[6] Jörg Benz, Hans Trachsel, and Ulrich Baumann. Crystal structure of the ATPase domain of translation initiation factor 4A from Saccharomyces cerevisiae – the prototype of the DEAD box protein family. *Structure*, 7(6):671–679, June 1999.

[7] Mark V Berjanskii and David S Wishart. The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Research*, 35(Web Server):W531–W537, May 2007.

[8] Jonathan M. Caruthers, Eric R Johnson, and David B McKay. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proceedings of the National Academy of Sciences*, 97(24):13080–13085, January 2000.

[9] Jeong Ho Chang, Yong Hyun Cho, Sun Young Sohn, Jung Min Choi, Ahreum Kim, Young Chang Kim, Sung Key Jang, and Yunje Cho. Crystal structure of the eIF4A-PDCD4 complex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3148–3153, March 2009.

[10] The UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 40(D1):D71–D75, January 2012.

[11] Frank Delaglio, Stephan Grzesiek, Geerten W Vuister, John Pfeifer, and Ad Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, November 1995.

[12] Roger Duncan, Susan C Milburn, and John W B Hershey. Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F. *The Journal of biological chemistry*, 262(1): 380–388, January 1987.

[13] Amr Fahmy and Gerhard Wagner. TreeDock: a tool for protein docking based on minimizing van der Waals energies. *Journal of the American Chemical Society*, 124(7): 1241–1250, February 2002.

[14] Amr Fahmy and Gerhard Wagner. Optimization of van der Waals energy for protein side-chain placement and design. *Biophysical journal*, 101(7):1690–1698, October 2011.

[15] Pinghui Feng, David N. Everly, and G. Sullivan Read. mRNA Decay during Herpes Simplex Virus (HSV) Infections: Protein-Protein Interactions Involving the HSV Virion Host Shutoff Protein and Translation Factors eIF4H and eIF4A. *Journal of Virology*, 79(15):9651–9664, August 2005.

[16] Dominique P Frueh, Alison Leed, Haribabu Arthanari, Alexander Koglin, Christopher T Walsh, and Gerhard Wagner. Time-shared HSQC-NOESY for accurate distance constraints measured at high-field in (15)N-(13)C-ILV methyl labeled proteins. *Journal of Biomolecular NMR*, 45(3):311–318, November 2009.

[17] Yuki Fujita, Masako Oe, Tatsuya Tutsumino, Shigenobu Morino, Hiroaki Imataka, Koji Tomoo, and Toshimasa Ishida. Domain-dependent interaction of eukaryotic initiation factor eIF4A for binding to middle and C-terminal domains of eIF4G. *Journal of Biochemistry*, 146(3):359–368, September 2009.

[18] Maayan Gal, Katherine A Edmonds, Alexander G Milbradt, Koh Takeuchi, and Gerhard Wagner. Speeding up direct (15)N detection: hCaN 2D NMR experiment. *Journal of Biomolecular NMR*, 51(4):497–504, December 2011.

[19] Kevin H Gardner and Lewis E Kay. Production and Incorporation of 15N, 13C, 2H (1H- 1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *Journal of the American Chemical Society*, 119(32):7599–7600, August 1997.

[20] Kevin H Gardner, Geoffrey Mueller, Randall Willis, and Lewis E Kay. A robust and cost-effective method for the production of Val, Leu, Ile (1) methyl-protonated 15N-

, 13C-, 2H-labeled proteins. *Journal of Biomolecular NMR*, 13(4):369–374, January 1999.

[21] TD Goddard and DG Kneller. SPARKY 3.

[22] Patrice Gouet, Emmanuel Courcelle, David I Stuart, and Frederic Metoz. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, 15(4):305–308, 1999.

[23] John D Gross, Nathan J Moerke, Tobias von der Haar, Alexey A Lugovskoy, Alan B Sachs, John E G McCarthy, and Gerhard Wagner. Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E. *Cell*, 115(6): 739–750, December 2003.

[24] Peter Güntert, C. Mumenthaler, and Kurt Wüthrich. Torsion angle dynamics for NMR structure calculation with the new program Dyana. *Journal of Molecular Biology*, 273 (1):283–298, 1997.

[25] Torsten Herrmann, Peter Güntert, and Kurt Wüthrich. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*, 319(1): 209–227, May 2002.

[26] Sven G Hyberts, Dominique P Frueh, Haribabu Arthanari, and Gerhard Wagner. FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. *Journal of Biomolecular NMR*, 45(3):283–294, August 2009.

[27] Sven G Hyberts, Koh Takeuchi, and Gerhard Wagner. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *Journal of the American Chemical Society*, 132(7):2145–2147, February 2010.

[28] Sven G Hyberts, Alexander G Milbradt, Haribabu Arthanari, and Gerhard Wagner. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *Journal of Biomolecular NMR*, pages 1–13, February 2012.

[29] Hiroaki Imataka and Nahum Sonenberg. Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol Cell Biol*, 17(12):6940–6947, December 1997.

[30] Eric R Johnson and David B McKay. Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase. *RNA*, 5(12):1526–1534, January 1999.

[31] Martin Karplus. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of the American Chemical Society*, 85(18):2870–2871, September 1963.

[32] Rochus L. J. Keller. *The Computer Aided Resonance Assignment Tutorial*. Cantina Verlag, Goldau, 2004. URL `http://www.nmr.ch`.

[33] Nadia L Korneeva, Eric A First, Clint A Benoit, and Robert E Rhoads. Interaction between the NH2-terminal domain of eIF4A and the central domain of eIF4G modulates RNA-stimulated ATPase activity. *The Journal of biological chemistry*, 280(3): 1872–1881, January 2005.

[34] Barry J Lamphear, Regina Kirchweger, Tim Skern, and Robert E Rhoads. Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation. *The Journal of biological chemistry*, 270(37):21975–21983, September 1995.

[35] Nicole Laronde-Leblanc, Arti N Santhanam, Alyson R Baker, Alexander Wlodawer, and Nancy H Colburn. Structural basis for inhibition of translation by the tumor suppressor Pdcd4. *Mol Cell Biol*, 27(1):147–156, January 2007.

[36] Roman A. Laskowski, J. Antoon C. Rullmann, Malcolm W. MacArthur, Robert Kaptein, and Janet M. Thornton. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, 8(4):477–486, January 1996.

[37] Portia G Loh, Hsin-Sheng Yang, Martin A Walsh, Qing Wang, Xiaoxing Wang, Zhihong Cheng, Dingxiang Liu, and Haiwei Song. Structural basis for translational inhibition by the tumour suppressor Pdcd4. *EMBO J*, 28(3):274–285, January 2009.

[38] Jon R Lorsch and Daniel Herschlag. The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry*, 37(8):2194–2206, February 1998.

[39] Jon R Lorsch and Daniel Herschlag. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry*, 37(8):2180–2193, February 1998.

[40] Joseph Marcotrigiano, Ivan B Lomakin, Nahum Sonenberg, Tatyana V Pestova, Christopher U T Hellen, and Stephen K Burley. A Conserved HEAT Domain within eIF4G Directs Assembly of the Translation Initiation Machinery. *Molecular Cell*, 7(1): 193–203, January 2001.

[41] Assen Marintchev and Gerhard Wagner. eIF4G and CBP80 Share a Common Origin and Similar Domain Organization: Implications for the Structure and Function of eIF4G. *Biochemistry*, 44(37):12265–12272, September 2005.

[42] Assen Marintchev, Katherine A Edmonds, Boriana Marintcheva, Elthea Hendrickson, Monika Oberer, Chikako Suzuki, Barbara Herdy, Nahum Sonenberg, and Gerhard Wagner. Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell*, 136(3):447–460, February 2009.

[43] Catherine Mazza, Alexandra Segref, Iain W Mattaj, and Stephen Cusack. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J*, 21(20):5548–5557, October 2002.

[44] Shigenobu Morino, Hiroaki Imataka, Yuri V Svitkin, Tatyana V Pestova, and Nahum Sonenberg. Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol Cell Biol*, 20(2): 468–477, January 2000.

[45] Monika Oberer, Assen Marintchev, and Gerhard Wagner. Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev*, 19(18):2212–2223, September 2005.

[46] Ali R Özeş, Kateryna Feoktistova, Brian C Avanzino, and Christopher S Fraser. Duplex unwinding and ATPase activities of the DEAD-box helicase eIF4A are coupled by eIF4G and eIF4B. *Journal of Molecular Biology*, 412(4):674–687, September 2011.

[47] Richard D Palmiter. Regulation of protein synthesis in chick oviduct. *Journal of Biological Chemistry*, 247(20):6450, 1972.

[48] Tatyana V Pestova, Christopher U T Hellen, and Ivan N Shatsky. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol*, 16(12):6859–6869, December 1996.

[49] George W. Rogers, Nancy J. Richter, Walt F. Lima, and William C Merrick. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. *Journal of Biological Chemistry*, 276(33):30914–30922, January 2001.

[50] Nadja Rozovsky, Aimee C. Butterworth, and Melissa J Moore. Interactions between eIF4AI and its accessory factors eIF4B and eIF4H. *RNA*, 14(10):2136–2148, January 2008.

[51] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. 2010.

[52] Patrick Schütz, Mario Bumann, Anselm Erich Oberholzer, Christoph Bieniossek, Hans Trachsel, Michael Altmann, and Ulrich Baumann. Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 105(28): 9564–9569, July 2008.

[53] Patrick Schütz, Tobias Karlberg, Susanne van den Berg, Ruairi Collins, Lari Lehtiö, Martin Högbom, Lovisa Holmberg-Schiavone, Wolfram Tempel, Hee-Won Park, Martin Hammarström, Martin Moche, Ann-Gerd Thorsell, and Herwig Schüler. Comparative structural analysis of human DEAD-box RNA helicases. *PloS one*, 5(9), 2010.

[54] Charles D Schwieters, John J. Kuszewski, Nico Tjandra, and G. Marius Clore. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, 160(1):65–73, January 2003.

[55] Charles D Schwieters, John J. Kuszewski, and G. Marius Clore. Using Xplor–NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 48(1):47–62, March 2006.

[56] Toru Sengoku, Osamu Nureki, Akira Nakamura, Satoru Kobayashi, and Shigeyuki Yokoyama. Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. *Cell*, 125(2):287–300, April 2006.

[57] Yang Shen, Frank Delaglio, Gabriel Cornilescu, and Ad Bax. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4):213–223, June 2009.

[58] Evangelos Sisamakis, Alessandro Valeri, Stanislav Kalinin, Paul J. Rothwell, and Claus A.M. Seidel. Chapter 18 - Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. In Nils G Walter, editor, *Methods in Enzymology*, pages 455–514. Academic Press, 2010.

[59] Randall M. Story, Hong Li, and John N. Abelson. Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proceedings of the National Academy of Sciences*, 98(4):1465–1470, January 2001.

[60] Chikako Suzuki, Robert G Garces, Katherine A Edmonds, Sebastian Hiller, Sven G Hyberts, Assen Marintchev, and Gerhard Wagner. PDCD4 inhibits translation initiation by binding to eIF4A using both its MA3 domains. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9):3274–3279, March 2008.

[61] Yuri V Svitkin, Arnim Pause, Ashkan Haghighat, Stephane Pyronnet, Gary Witherell, Graham J Belsham, and Nahum Sonenberg. The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA*, 7(3):382–394, January 2001.

[62] N. Kyle Tanner, Olivier Cordin, Josette Banroques, Monique Doère, and Patrick Linder. The Q Motif: A Newly Identified Motif in DEAD Box Helicases May Regulate ATP Binding and Hydrolysis. *Molecular Cell*, 11(1):127–138, January 2003.

[63] Quincy Teng. *Structural biology: practical NMR applications*, page 93. Springer Verlag, 2005.

[64] Geerten W Vuister and Ad Bax. Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNH. alpha.) coupling constants in 15N-enriched proteins. *Journal of the American Chemical Society*, 115(17):7772–7777, 1993.

[65] Andy C Wang and Ad Bax. Determination of the Backbone Dihedral Angles in Human Ubiquitin from Reparametrized Empirical Karplus Equations. *Journal of the American Chemical Society*, 118(10):2483–2494, January 1996.

[66] David S Waugh. Genetic tools for selective labeling of proteins with alpha-15N-amino acids. *Journal of Biomolecular NMR*, 8(2), September 1996.

[67] David S Wishart, Brian D Sykes, and Frederic M Richards. Relationship Between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure. *Journal of Molecular Biology*, 1991.

[68] Toshio Yamazaki, Julie D Forman-Kay, and Lewis E Kay. Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *Journal of the American Chemical Society*, 23(115):11054–11055, November 1993.

[69] Hsin-Sheng Yang, Aaron P. Jansen, Anton A. Komar, Xiaojing Zheng, William C Merrick, Sylvain Costes, Stephen J. Lockett, Nahum Sonenberg, and Nancy H Colburn. The Transformation Suppressor Pdcd4 Is a Novel Eukaryotic Translation Initiation Factor 4A Binding Protein That Inhibits Translation. *Molecular and Cellular Biology*, 23(1):26–37, January 2003.

[70] Hsin-Sheng Yang, Myung-Haing Cho, Halina Zakowicz, Glenn Hegamyer, Nahum Sonenberg, and Nancy H Colburn. A Novel Function of the MA-3 Domains in Transformation and Translation Suppressor Pdcd4 Is Essential for Its Binding to Eukaryotic Translation Initiation Factor 4A. *Molecular and Cellular Biology*, 24(9):3894–3906, January 2004.

[71] Gulnara Zh Yusupova, Marat M Yusupov, J H D Cate, and Harry F Noller. The Path of Messenger RNA through the Ribosome. *Cell*, 106(2):233–241, July 2001.

[72] Pei Zhou, Alexey A Lugovskoy, and Gerhard Wagner. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *Journal of Biomolecular NMR*, 20(1):11–14, 2001.

Colophon

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. A template that can be used to format a PhD thesis with this look and feel is freely available online at `https://github.com/suchow/`.